

nature

April 22, 1976

Don't sit on the milestone

It is with some sense of relief, even pleasure, that it can be reported that the United States and the Soviet Union have actually managed to conclude an agreement on limiting underground nuclear tests to yields of less than 150 kilotons. Two months ago we were gloomily predicting that the superpowers would not meet their modest deadline of March 31, and we were wrong. In the process of negotiations it looks as if some real headway has been made towards establishing the principle of on-site inspection so important to the Americans, worried about cheating, and so tiresome to the Russians, worried about espionage. It has been agreed that the explosives for peaceful purposes shall not come in quantities of greater than 150 kilotons separately, although several may be fired simultaneously. Inspectors will monitor adherence, presumably by some down-hole device attached to each explosive to measure yield.

The satisfaction is that the threshold treaty, despite everything, is intact and the superpowers have done what they said they would. The vigour of the Soviet Plowshare programme, in which projects as big as the diversion of rivers still feature, has always been a worry to those who saw a treaty and a peaceful programme as incompatible for arms-control purposes. The neat formula by which this worry has been resolved offers some hope for future developments within the same framework.

This week's Geneva conference on test-ban monitoring will no doubt grudgingly acknowledge the Soviet-American agreement, but participants will also go on to say, privately if not in public, that

- Bilateralism is too cosy
- 150 kilotons is too high
- efforts to lower the threshold must go on.

Both superpowers have been vigorous in their high yield weapon-testing in the run-up to the treaty. In the United States advanced development research and testing has been put off for the last year and a half, according to ERDA's congressional supporters, while work was directed towards preparing for the treaty. There is thus a great danger that now that a milestone has been reached, the superpowers will sit on it. What better than that the Swedes, the Japanese, the Canadians and all other countries which have tried for years to make the superpowers see that vertical proliferation is as offensive as horizontal proliferation, should urge a painless formula gradually to reduce this threshold to zero. Such as:

$$\text{Threshold} = 150(1 - (\text{Year} - 1976)/10) \text{ kilotons.}$$

And how about some positive support for their efforts from the United Kingdom which has too much in recent years adopted a don't-rock-the-superpowers'-boat attitude engendered by nuclear subservience to the United States?

Mixture as before

ANYONE who thought that, because of the prominence given to the fight for the Labour leadership, the consequences of Harold Wilson's departure might entail a different if not more favourable orientation in the UK government's disposition towards science, can be forgiven for feeling frustrated. Under James Callaghan the government has already twice indicated that, in this respect at least, continuity (meaning no change) is the order of the day.

The first, crucial indication came, not with Denis Healey's Budget, but in a White Paper released the same day entitled *Cash Limits on Public Expenditure* (Cmd. 6440, HMSO). The limits, which apply to more than three-quarters of central government spending not counting social security benefits, are controls on the actual cash that may be spent in the year 1976-77 by specified blocks within government departments. The system supersedes the traditional budget estimates, which subsequent supplementary estimates used to inflate to compensate for escalated prices and costs during the year; now, if prices rise too high, the White Paper says, government purchases would have to be cut back.

The figures make the position clearer. The cash limit on Department of Education and Science's Science Budget, for example, which finances the activity of the Research Councils, is £226.1 million. The voted figure, announced earlier this year, is £215.949 million, and this remains the total available to the Research Councils. The

difference represents an amount which could conceivably be available if times became really hard and the needy Research Councils could present a convincing case, to both the Minister and the Treasury, for the created margin of flexibility to operate in their favour. The Department of Energy's expenditure on nuclear energy, to take another example, has a cash limit of £124.4 million, and the same strictures will apply.

What is less clear is how the cash limits are arrived at. They obviously incorporate a government estimate of the rate of inflation for the year—an estimate that is still (temporarily?) secret. But if the declared aim of 10% by the end of 1976 is not achieved, research backed by the councils may not go ahead. And, given the need to contain public expenditure, if the rate ends up lower than forecast, the cash limits could conceivably be reduced.

The other, less crucial indication of a no-change policy, came with Mr Callaghan's Cabinet reshuffle. A glance tells all: Fred Mulley stays at Education and Science, Tony Benn stays at Energy, Eric Varley stays at Industry, Roy Mason stays at Defence (the only major department which has just a single block—of £5,835 million—covering all its expenditure). Peter Shore moves to the mis-named Environment giant.

It is early days yet, but with junior positions unlikely to make much difference, science policy looks for the moment like staying stagnant at the top. □



Keeping our heads about our heads

Controversial questions are currently being raised about psychosurgery.

Keith Oatley of the Laboratory of Experimental Psychology at the University of Sussex comments on some of the issues

IN THE last few weeks public notice has been drawn to the use of physical and surgical interventions in the treatment of "mental illness". The release of the film "One Flew Over the Cuckoo's Nest", in which both electro-convulsive treatment and lobotomy were portrayed as techniques of punishment and control, was followed by a television documentary of a woman who had bouts of violent behaviour, and who was operated on by a neurosurgeon to remove part of her hypothalamus. The results were then reported of an Inquiry which upheld allegations made at a "typical" British mental hospital that electroconvulsive treatment was given "casually" and on "a disturbingly large number" of occasions. All this comes at a time when the Royal College of Psychiatrists is planning a controlled study of psychosurgery on 200 patients in the London area.

Under these circumstances it seems appropriate to consider some important issues raised by psychosurgery: the attempt to control entirely mental and behavioural symptoms (with no organic pathology) by physical stimulation or lesioning of the brain. Two fairly extreme positions can be taken up, both with a good measure of enthusiasm, righteous indignation, and often some reluctance to examine the evidence.

One position is that psychosurgery is barbaric, infringes patients' rights, and produces irreversible damage to the brain (how can less brain be better than more?). What justification is there in any case, it is asked, for supposing that people who have difficulties in living are "ill" or need to have some part of their brain stimulated or removed?

At the other extreme some members of the medical profession claim that psychiatric disturbances are illnesses with physical causes, that it has been shown that for some of these, intervention in the brain produces alleviation of symptoms, and even cures, and that doctors are in a position both to recommend, carry out and evaluate such treatment. (This latter point for instance was the conclusion of *The Times* medical correspondent Dr Tony Smith. "Only another doctor can be certain that a line of medical treatment was inadvisable, unethical, or simply wrong".)

Neither position squares well either with our physiological, psychological or sociological understandings of psychologically disturbed, or disturbing behaviour. Three kinds of consideration seem to be most relevant. First, there are the reported results of psychosurgical procedures which have already been undertaken. Secondly, there is evidence from experimental animal work (from some of which the psychosurgical techniques were originally derived), and from experimental analysis of, for example, accidental brain damage in people. Finally, there are some fundamental considerations of how people, including doctors and patients, ought to behave towards one another, backed if necessary with legal sanctions. It is worth considering one illustrative example from each of these.

In a recent issue of the *British Journal of Psychiatry* (128, 226-240; 1976), N. Mitchell-Heggs, D. Kelly and A. Richardson of St George's Hospital Medical School report on 66 patients given psychosurgery, and followed up over a mean period of 16 months. The operations were multifocal. Lesion-making electrodes were placed stereotactically in 10 sites of each brain (typically three in each frontal lobe, and two in each cingulum). The thermo-coagulative lesions were aimed at various limbic tracts, and each estimated at 6 mm in diameter. Before and after operation patients were assessed on an impressive number of scales and inventories, and in 28 cases by a semi-independent assessor (a retired psychiatrist). The report was that 73% of the patients were improved at 6 weeks and 76% at 16 months. The paper is excellently detailed by the usual standards of such reports. The authors take pains to discuss the prerequisites for the operation (that is, an exhaustive trial of less radical treatments), and give a lot of statistics about their assessments.

Various comments can however be made. There is no control group, for example, and with the best will in the world the assessment is done by people who have a strong interest in reporting success of the procedure. Also, although one gets the impression that many patients did benefit substantially from the treatment (various remarks being made about "returning to work" and so on), the measures of improvement did not include assessment of

whether the patients were improved with respect to their own objectives in life, or with respect to whether they were easier to contain in institutional care. Moreover, many of the scales used were heavily biased. Clinical rating scales, for instance, had three categories for "improved", one for "unchanged" and one for "worse". It is known that to approximate to fairness, rating scales should be symmetrical about the null hypothesis of "no change".

Regarding the experimental evidence (mostly from animals), this indicates that brain anatomy is no less variable than the anatomy of human faces, that stereotaxic placement of electrodes is in any case subject to errors (for example, of electrodes bending or being misdirected and thus ending up in places other than those planned). Even disregarding these there is no evidence that parts of the brain are related in a one-to-one fashion or any other simple way to categories of behaviour such as aggression or obsessiveness. (The best review of these matters in the context of psychosurgery is by E. Valenstein, *Brain Control* (Wiley; 1973)).

Finally, although careful consideration of the experimental evidence from animal work would dissuade one from attempting psychosurgery, it does already take place, and some doctors believe it benefits some extremely disabled people considerably. If psychosurgery is to take place at all, what we do need is a proper trial, with proper controls, with informed consent of all subjects, and the participation of people (other than medics.) who know how to design and assess such experiments.

The trial should compare a group of patients who have been referred for psychosurgery (on the basis of severe disablement and failure of previous treatments). Half should have the psychosurgery, and the other half a further attempt at the best available non-surgical treatment (perhaps a combination of chemotherapy and behaviour therapy). All participants should be given supportive psychotherapy and rehabilitation help throughout. Assessment before and after treatment should be undertaken by several independent people, of different psychiatric persuasions, and without knowledge of what treatment each patient is given.

There are ethical problems; but these are less questionable than the current practice of doctors carrying out research without informed consent of all subjects in ill controlled studies which continue to leave room for doubt about the efficacy of procedures which then continue to be practised. □

Science and the media

David Fishlock, Science Editor of the *Financial Times*, looks at two recent publications* on a controversial relationship

SOME YEARS ago, while writing a science feature for a daily newspaper, I was advised by one scientist to ring up another for help on one specific point. But the professor in question, unknown to me, was not willing to talk on the 'phone and suggested I write. I said—perhaps impatiently—that I wanted to develop and verify the point right away and “can't wait three months for a reply”. The professor would not budge but did ring my Editor to complain about my “rudeness”.

A year or two later I met the professor at a Press conference called by her university to discuss the results of Government-sponsored work in which she had been closely involved. On that occasion she was helpful on everything connected with the Government-funded part of the work, but could not be drawn at all on the (very considerable) commercial implications of her university research. It turned out that she was also advising a large private company.

The readiness with which this university scientist was willing to place part of her knowledge and experience beyond the reach of the public (including her fellow scientists) is a good illustration of the scope for confusion about the respective roles of scientist and journalist. For those scientists who are interested in what the Press and broadcasting channels do with science—and there can be few research workers in these straitened times so unworldly as to be indifferent to the media's activities—these two short books, of identical title, offer a valuable grounding. One is the report of a joint scientist-journalist study group set up by the British Association (BA); the other, a contribution to OUP's excellent Science and Engineering Policy series.

The BA report deals specifically with the quality of the presentation of science subjects to a wider public, improvement in which the BA believes calls for “a much closer relationship between scientists and journalists than exists generally today”. It acknowledges that “several factors limit the frankness between scientist and journalist”, among them contractual obligations and fears that the informa-

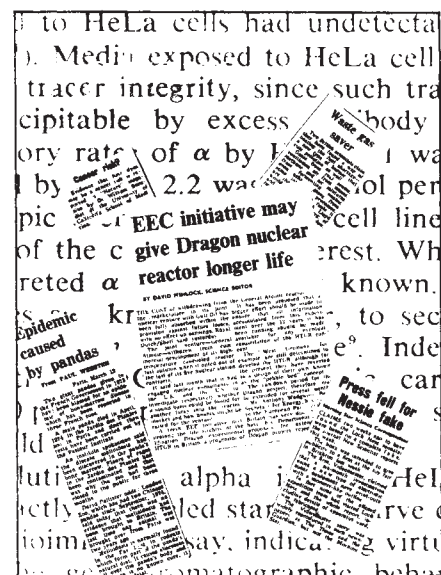
tion may be over-dramatised. “It is not that he mistrusts the good science correspondent, who can be expected to present issues in a fair and balanced way, but that he fears any airing in the media runs the risk that the story will be taken up, sensationalised and distorted by others”.

In 16 pages the BA report sets out very lucidly what the science journalist or broadcaster is trying to do, and why, what handicaps he must work under, and why he is so dependent upon the cooperation of scientists—and good scientists to boot. It has a salutary word of warning about “scientists, mostly with a gift for entertaining the public, who are tempted to exploit the media for personal reasons”, and whose ability to dazzle a non-scientific audience may lead them to make “wild and reckless statements”. It advises the media to be as wary of such scientists as they are of political adventurers. The corollary of course, is that if the genuine scientist will not help the journalist, he may well be tempted to make greater use of charlatans.

The BA study also touches upon the so-called growing public disenchantment with science, evidence for which it finds is thin. But it suggests that some journalists who believe in it allow their writing to reflect hostility towards science, and thus help to foster hostility between the public and science.

Dr Peter Farago's book covers broadly the same territory at about five times the length. The author is a scientist who edits technical journals and who thus has a toe in both camps. He tends to take a gloomier view of relations between scientists and journalists than the BA working party, which seems to reflect the views of those science correspondents on whom he has relied for first-hand experience of the craft. “The need for scientists to explain was a constant accompaniment to all discussions, together with the hurt accusation that they will not be properly questioned by reporters”.

Farago notes that only one newspaper in Britain carries a daily item on science. This is the *Nature-Times* Science Report, modelled on the paper's famous Law Reports, from which I once culled a painfully funny item entitled “Raving and gobbling”. There is little danger, however, of anyone doing himself an injury while reading one of its science reports, which are workmanlike but quite uninspired precis of, usually, original scientific papers, tucked away in a “ghetto” at



the back of the paper and the bottom of the page.

The *Nature-Times* Science Report is a good illustration of some of the major faults and failings of science writing and attitudes towards it today. It is “ghetto journalism” inasmuch as it caters for a select group of readers. Farago found no one on newspapers who found the reports readable—a strong condemnation of any regular newspaper item. They cry out for the touch of a sympathetic editor who knows how to write a good headline, how to inject some flair into the flat prose—all of which can be done, I should assure the purists, without distorting the facts.

But all the skills of the most sympathetic sub-editor could not overcome the handicap of having a guaranteed if inconspicuous place in the paper each day, which insulates it from the need to compete for space and attention. As a consequence it becomes isolated from the substance of all other reporting, no longer part of the fabric of everyday news. This may even redound in attitudes towards other science stories which get less attention because the paper thinks it is already doing its duty towards science by allocating a corner each day.

Neither of these two studies reach any very profound conclusions. They agree that science writing can be very good but often is not, and that it would probably be in everyone's interest if scientists and those who write about their work were to take a more sympathetic view of each other's activities. The seven conclusions set out in the BA report offer sensible advice for improving relations, and include the idea of more closely studying just what the public knows and thinks about science and where its ideas are coming from. □

**Science and the Media*. (British Association Publication 76/1: London, March 1976).

Science and the Media. (Science and Engineering Policy Series.) By Peter Farago. Pp. 95. (Oxford University: London, April 1976).

WEST GERMANY

WEST GERMANY'S Economics Minister, Dr Hans Friderichs, recently announced that Bonn and Moscow had scrapped joint plans, first mooted some 18 months ago when Chancellor Helmut Schmidt visited Moscow, for the construction of a huge nuclear power station at Kaliningrad (formerly Königsberg) on the Soviet-Polish border.

Under the customary sort of no-cash-involved deal favoured by the USSR, Kraftwerk Union and other West German firms would have built the reactor with West German money (reportedly over £300 million), and given the USSR both the technology and the means of supplying electricity to its industrialised west; the USSR would have paid in the form of deliveries of electricity to West Berlin, which is surrounded by East German territory.

The reason given by Dr Friderichs for the cancellation concerned the USSR's proposals for pricing the electricity, which West Germany's electricity companies regarded as unacceptably high. But the suggestion that the project be abandoned came from the USSR when Dr Friderichs visited Moscow, and it is clear that the high prices sought by the Soviet Union in fact merely ensured the collapse of a deal to which East Germany had taken strong exception because of the inclusion of West Berlin.

The East Germans objected to the idea of greater security for West Berlin implied by the supply of electricity to the city. West Berlin may now have to construct conventional plants, although there is a plan for an underground nuclear reactor, while the Soviet Union, for the moment at least, will not obtain its 1300 MW power station.

● The recent State visit to West Germany by Egypt's President Sadat, widely regarded as further cementing relations between the two countries, included a highly publicised excursion to Germany's large nuclear power plant at Biblis. The possibility that Egypt will be a purchaser of German nuclear technology is not being discounted, particularly as Egypt's Minister for Science, Research and Nuclear Energy, El Guebeily, when he visited Germany at the end of January, was accompanied by the President of the Egyptian Academy for Research and Technology and the Deputy Director of the Egyptian Atomic Energy Authority.

Contracts arranged under the German-Brazilian agreement on nuclear cooperation of June 1975 have meanwhile been signed. They provide for uranium production, processing and enrichment as well as for the construc-

tion of nuclear reactors and of a re-processing plant.

● Closely following publication by the European Commission of its second environmental programme for 1977 to 1981, in which priority is given to reductions in water pollution as well as reductions in waste and noise abate-



ment, a group of West German environment experts has argued in a report presented to the Bonn government that a clean-up of the Rhine river is "primarily a German task" because the greatest share of the pollution is of German origin.

The group also suggests that the cost is "not too high" at about £300 million a year, arguing that the Federal government could bear one-third of this, state and local government could finance another third, while the industries discharging effluent into the river would pay for the remainder. The group recommends an effluent tax for industry to encourage Rhine-based companies to purify their own waste.

The report itself says thermal pollution and heavy metal pollution are not problems, but it does say that Rhine pollution has worsened, with salt and oxygen content often at unacceptable levels. In the week the report was published, however, it was announced from Paris that the environment ministers of Germany, France, Holland, Luxembourg and Switzerland (countries drained by the Rhine and its tributaries) had agreed on steps to reduce the excessive levels of salt.

The proposed scheme, which would start in two years if final agreement is reached, would cost about £6 million, but the ministers were unable to concur on how the running costs (around £450,000 a year) should be shared.

● "To strengthen the reciprocal information and communication process between State and Industry in the sphere of practical promotion of research and technology, and to offer

starting points for a constructive critique of State measures". That is the aim of a document on state-aided research, entitled *Förderfibel* ("promotion primer"), published recently by the German Ministry of Research and Technology and drawn up by the Institut für Systemtechnik und Innovationsforschung (ISI). Directed particularly to small and medium size companies and to groups of researchers and inventors, the document has a preface by the Minister, Hans Matthöfer, which sets out a recommended procedure for budding inventors and research workers seeking support. The idea is to reduce the waste of time both by themselves and by officials they will approach. It is the ministry's first-ever detailed guide to government support for research.

● The chemical cycle is turning upwards in West Germany. The country's three major chemical concerns—Bayer, BASF and Hoechst—have all now reported a pick-up in business, matching exactly a similar recovery in Britain. As in Britain, the industry has been through a very depressed period, so much so that the upward trend in domestic trade will need to continue and begin revealing itself in foreign trade too if overall activity, including research, is to attain the sort of levels to which this huge and important sector of German industry has grown accustomed.

● Scientists, as much as anyone else in West Germany's academic community, are concerned about the implications of the tradition known as *Berufsverbote*, under which the political reliability of people can be tested before they are admitted into the service of the state. Among the best documented cases in respect of higher education, most involve teachers in the social sciences. Estimates of the overall number of cases of refusal of appointment in the Civil Service go into the hundreds, and the number of screenings into hundreds of thousands.

The academics' concern is understandable: not only does it involve questions of basic human rights; it also affects them directly because they, like members of the regular public administration, are civil servants. The idea behind the practice is to prevent "extremists" taking up public appointments, but the major parties disagree less on the idea itself than on whether debarment should be determined merely by a person's membership of a party hostile to the Constitution rather than the particulars of each individual case. The difference has prevented agreement in the Upper House on the so-called Extremist Bill.

ISLAMIC SCIENCE

Scientists gather in Riyadh

Some two hundred Muslim scientists met at the University of Riyadh in Saudi Arabia last month for the Islamic Solidarity Conference in Science and Technology. Michel Chodkiewicz of La Recherche reports

A SCIENTIFIC renaissance in Muslim countries is one thing. An Islamic scientific renaissance is another. The former has begun, more or less successfully, in several countries; the latter remains to be done . . . if it has any meaning at all. For many Westerners, and for some Muslims this last point is far from being proved.

This doubt was manifestly not shared either by those scientists present in Riyadh for the "Islamic solidarity conference in science and technology", or by the Saudi authorities who had bestowed munificent patronage on the meeting. Several Saudi ministers participated and King Khaled presided at the opening session. For all of these there was no conflict between Islam and science, either in the principles ("Seek out science, be it as far away as China", the Prophet had said) or in the facts (the brilliance of the abbasid era was witness to that).

But complacent contemplation of past glories must give way now to serious planning for the future; the majority of the presentations and discussion during the conference were very down to earth. From problems of nuclear energy, naval construction and scientific documentation to those of higher education, the subjects tackled seemed to have two priorities: that of a better balance within the Islamic world between financial resources and manpower; and that of a very powerful drive for research and development, without which the Muslim world will never reach the "critical mass" above which it will be taken seriously scienti-

fically, economically and politically.

The first of these problems—that of finance and manpower—is particularly evident in Saudi Arabia with its crying lack of manpower. (During our stay the government announced that civil servants were henceforth permitted—even encouraged—to work for private enterprise in their spare time.) One of the unofficial aims of the conference was to profit by this gathering of scientists and to organise a reverse brain-drain. Amongst the 185 Muslim scientists present, 27 worked in the USA, and it was particularly to them that a call was addressed to come and pursue their work in the Islam world. But researchers or technicians working in the resource-poor manpower-rich Islamic countries (such as Pakistan and Egypt) were also tempted by the salaries offered in the oil states.

In the long run this brain-drain, internal to the Islam world, may well be dangerous. The balance between the "umma" interest (the Islamic community as a whole) and those often divergent interests of constituent states will be difficult to get right. Moreover, the scientists working in the non-Islamic world will only return to a satisfactory environment, stimulating professional contacts and so on. The creation of centres of excellence could help resolve this problem: to this end, a first step will be taken quite quickly, thanks to the founding of an Islamic Foundation for Science and Technology, probably at Taif, the Arabian city with the most pleasant climate, which has the added advantage of being near both Jedda, where a huge university is being built, and Mecca, site of another planned university.

On most aspects of the arrangements between Islamic countries, visitors, not representing their countries officially, could only express wishes. However, the creation of a permanent Conference Secretariat, entrusted to the direction

of the Rector of Riyadh University, will doubtless lead to some concrete action in the fields of, for instance, of documentation and scientific information (unifying the technical vocabulary, publishing scientific periodicals, organising seminars and symposia) and in collation of statistical data needed for planning.

Amongst the Conference's recommendations was the establishment of a consultative body comprising scientists and engineers working in military research and armaments. This initiative gave further substance to the already announced decision that several Arab countries had begun to develop a military-industrial community in Egypt. But to move out in this field from the Arab to the Islamic world will not be without its problems. It is worth noting that in his shrewd and much applauded speech at the final assembly Professor Fazlollah Reza, who was at the conference as a physicist but who is also Iranian ambassador to Canada, made remarks on all the resolutions adopted . . . with the exception of that concerning military consultation, which he passed over in silence.

● **Salah Galal writes:** A total of nine "Scientific Sessions" discussed specific issues. These included natural resources; industrial planning and growth; light and heavy industries and, specifically, chemical industries and manufacturing processes; university education, including its responsibilities towards industry, and technical education and training; research in universities and research centres; and translation and documentation.

It was three special "Projects Committees" which proposed the establishment of the Islamic Foundation for Science and Technology and the establishment of an Islamic Institution for the development of manpower, and produced the recommendations concerning armaments and military research.

RECYCLING

Paper chase

A recent publication from Friends of the Earth, *Ecological Paper Buying*, featuring a variety of recycled paper, urges the extensive substitution of waste paper for pulp as the raw material for paper making, in order to reduce Britain's huge annual pulp import bill. More waste paper, it says, could be used in newspapers, writing and printing papers and tissues, which currently take up 45% of the market.

The FOE spotlight a vicious circle, in which there is little demand for recycled paper because it is not widely available or insufficiently publicised, but not more widely available because there is little demand. In fact the matter is more complicated than that. As the Waste Management Advisory Council in the UK said in its first report last January, there would have to be an increase of 50% in the usage of waste paper by the early 1980s (compared to 1973–74) to save £100

million a year on imports given existing technology. The value of those imports was £270 million in 1974.

That means an extra 1.1 million tons of waste paper a year, which has to be collected and stored. The problems this presents, coupled with the problems of cyclical demand, seem to make a buffer stock necessary. One estimate of the finance involved in an excess stocks scheme is £10 million—a sum that is apparently not around at the moment.

correspondence

Genetic engineering

SIR,—In recent months, molecular biologists have discussed the dangers which can arise from careless or ill-advised experiments in genetic engineering, and have attempted, at various meetings, to lay down guidelines as safeguards against these dangers. Since experiments in this field will, for some years to come, be possible only at specialised centres associated with drawing up these guidelines, their voluntary observation provides a degree of protection against dangerous experiments. The reports of such conferences, however, suggest an air of complacency, doubtless unintentional, giving the layman, or the scientist in other disciplines, an impression that all is under control. This is dangerous.

Experience shows that it would be utopian to expect that such safeguards will continue to be accepted by all workers in the field. If it appears possible to use genetic engineering to cause damage or disruption for military purposes (several such possibilities can be envisaged), there will be organisations and states willing to sponsor and scientists willing to undertake such research, motivated by intellectual curiosity, or mistaken patriotism, or even a desire for gain and power. Under the circumstances of secrecy which would almost certainly be imposed, voluntary controls would be ineffective.

Fortunately secrecy in research is rarely absolute and is difficult to maintain in the face of informed vigilance. If the dangers of biological warfare appear to have receded in recent years, it is largely because of the vigilance of organisations like SIPRI, and the discussions stimulated by their reports. When everyone can recognise a danger and can work on counter measures, the danger becomes less acute.

It would thus appear to be an opportune time for molecular biologists to set up their own vigilance machinery, not as a police activity which would be repugnant to most scientists, but as information groups, which collect information on all known ongoing research, and can discuss and publicise the implications and the dangers of such work. Periodical publication of such reports in journals of general scientific interest would create an awareness and interest in these problems, and the publicity might discourage many persons who might

otherwise be tempted into clandestine research.

In the present situation, the scientists' only safeguard against misuse of knowledge is the development of informed public opinion. Creation of such opinion should therefore be accepted as a responsibility of those who are exploring newer fields of research.

Yours faithfully,

A. N. D. NANAVATI

Bombay Natural History Society,
Bombay,
India

Superstar technologies

SIR—I am very pleased that you considered our report on "Superstar Technologies" worthy of serious attention. Such constructive criticism as yours is important for any subsequent discussion of the report and its recommendations, and also helps improve our work.

The main point at issue between us is this: does technology, like science, need peer-review and mutual criticism? We think so; you doubt it. What is our alternative? Apparently not very much, since you are suspicious of any serious restraint (even if only moral) on a technologist's freedom of action, lest he emigrate or his enthusiasm be snuffed out.

The working party was well aware of this point; you will find a reference (No. 27) to the paper by Freeman Dyson where it is argued eloquently. But they and the Council felt that the scope and power of modern technologies are so enormous, particularly in their political, economic and social factors, that self-monitoring by enthusiastic experts is simply not good enough. Hence our introducing the idea of "Superstar Technologies" and hence our advocacy of a Technology Implications Commission with its carefully delineated functions.

Yours faithfully,

MICHAEL SWANN

Council for Science and Society,
St Andrews Hill,
London, UK

Isolative sound-change

SIR,—It may be suggested that the most important unsolved problem of Historical Linguistics is: why do isolative sound-changes take place? The nature of isolative sound-change is best explained by an example. Anglo-Saxon *bāt* changed into Middle English *bōt*

(to rhyme with *fort*), which changed into Modern English *boat*; Anglo-Saxon *hām* changed into Middle English *hōm*, which changed into Modern English *home*. It is possible to add very many examples with Anglo-Saxon *ā* to these two, and, on the basis of these examples, we formulate the sound-changes "Anglo-Saxon *ā* changes to Middle English *ō*, which changes to Modern English [ou]" (where [ou] is the vowel-sound in *boat* and *home*, spelt in different ways). This change is not affected by surrounding consonants—some changes are—that is why it is called "isolative".

The reason for sound-changes of this nature is entirely unknown. In fact, rather few suggestions as to a reason have been made; there was, perhaps, more interest in the subject in the nineteenth century than there is to-day. It may be that the problem, if it ever is solved, will be solved from disciplines other than Historical Linguistics—Psychology comes to mind. It seems that self-analysis is unlikely to help. I have an example of an isolative sound-change in my own speech: I pronounce the word *fire* identically with *far* (and so do some other people), whereas many people pronounce it almost to rhyme with *buyer*. And so for all the rhyme-words (*tyre*, *mire*, etc.). This change—[aiə] to [a:] in phonetic terms—has certainly taken place in my speech during my lifetime. But I am not able to say why.

Yours faithfully,

ALAN S. C. ROSS

Southwick,
Sussex, UK



A hundred years ago

THE Rhind Lectures on Archaeology, in connection with the Society of Antiquaries of Scotland, will be given by Dr Arthur Mitchell, commencing on Tuesday last, and continued on the following Fridays and Tuesdays. There will be six in all, and the subject is, "Do we possess the means of determining scientifically the condition of Primæval Man and his Age on the Earth?" from *Nature*, 13, 495; April 20, 1876

news and views

Apollo asteroid Helin

from David W. Hughes

CONSIDERABLE excitement was caused by the discovery in January of asteroid 1976 AA because it has subsequently been found to be the first asteroid whose mean distance from the Sun is less than that of the Earth. Eleanor Helin (California Institute of Technology) discovered the object on photographic plates taken with the 18 inch Schmidt telescope of the Palomar Observatory on January 7. It was then a faint 14th magnitude object in the constellation of Gemini. Using 46 observations taken between January 7 and February 3 Brian Marsden (Smithsonian Astrophysical Observatory) calculated the orbit of the asteroid and this is shown in Fig. 1.

On page 691 of this issue of *Nature*, Morrison, Gradie and Rieke of the Lunar and Planetary Laboratory of the University of Arizona report that the asteroid has a mean diameter of 900 ± 200 m and a visual geometric albedo of 0.20 ± 0.07 . So Helin becomes the smallest asteroid ever to have its diameter measured accurately. (The asteroid will not be spherical but will have an irregular shape rather like that of Phobos, a satellite of Mars.)

A radiometric technique was used to obtain the diameter and albedo. The principle of the technique is simple. A fraction A (known as the albedo) of the total sunlight striking the body is reflected whereas the rest is absorbed. The reflected radiation appears as the visual brightness, whereas the energy absorbed by the body (assuming the

body is in thermal equilibrium and non-volatile) is re-emitted as thermal infrared radiation. The ratio between the reflected and re-emitted radiation gives the albedo. Combining this albedo with either the visual or the infrared brightness of the object gives the radius. Unfortunately there are a few snags. All the infrared radiation cannot be measured from ground-based observatories as the $8\text{--}14\text{ }\mu\text{m}$ and $17\text{--}28\text{ }\mu\text{m}$ windows of the Earth's atmosphere must be used. Second, the infrared emissivity of the soils and rocks on the small airless asteroid must be assumed as must also the dependence of the albedo on the solar zenith angle and on the wavelength of the incident energy. The rotation of the asteroid causes the unilluminated hemisphere to radiate some part of the absorbed solar radiation at the expense of the radiation from the illuminated hemisphere. The mean temperature of the observed disk is thus decreased by an amount dependent on the rotation period and the thermal inertia of the surface, leading to an over estimation of the radius. Morrison (*Icarus*, **19**, 1; 1973) finds that at wavelengths greater than about $10\text{ }\mu\text{m}$ the flux is sensitive to rotation and that all the above factors lead to a 10% uncertainty in the radius and a 20% uncertainty in the albedo.

Spectrophotometric studies indicate that asteroid Helin has a surface which resembles that of stoney, siliceous meteorites and not a carbonaceous chondrite as previously reported (*Sky*

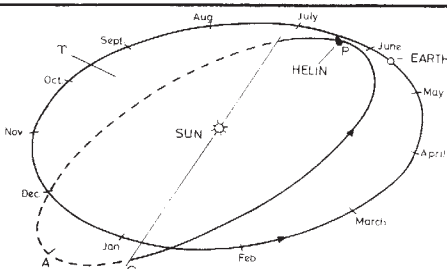


Fig. 1 The orbit of Helin (*IAC Circular*, No. 2915, February 24, 1976). Helin is shown at perihelion (P) and the position of the Earth at this time is also indicated. Helin will reach the aphelion of its orbit (A) on November 9 this year. (γ is the first point of Aries.)

and Telescope, **50**, 158; 1976). Siliceous surfaces seem to be common among Apollo and Amor asteroids.

Helin is the 20th Apollo asteroid so far discovered and joins her more distinguished cousins Icarus, Geographos and Toro in this group of asteroids characterised by having perihelia inside the Earth's orbit.

Shoemaker (CalTech) points out that asteroid Helin is also one of the easiest places to get to in our Solar System apart from the Moon. It makes a most attractive target for an unmanned space mission. Asteroids—half way in mass between meteorites and planets—probably are typical of the building blocks which coalesced to form the planets. Not suffering from convective differentiation and geophysical and atmospheric effects as the planets did, they are most probably in the form of the original condensate from the pre-Solar System nebula and as such furnish an outstanding clue to the origin of the Solar System. □

Plasmid exclusion and incompatibility

from J. R. Saunders

MANY bacteria carry self-transmissible plasmids that characteristically inhibit infection with the same or related plasmids. This phenomenon of superinfection immunity is the result of the presence of two genetically and physiologically distinct barriers, surface (entry) exclusion and incompatibility.

Surface exclusion is probably a device to prevent unnecessary mating within a population of homologous male cells. Exponential cultures of bacteria carrying a transmissible plasmid are therefore poor recipients in matings with donor cells carrying a homologous sex factor. Exclusion is believed

to involve recognition by the recipient of a specific (homologous) sex pilus on the donor and result in blockage of the transfer of DNA. It is mediated, at least in part, by the *traS* gene which forms part of the transfer operon of the F factor (Helmuth and Achtman, *Nature*, **257**, 652; 1975; Willetts, *J. Bact.*, **118**, 778; 1974). But the *traS* product is required only in the recipient and not the donor for surface exclusion to occur. The observation that disruption of the mucopolysaccharide layer of donor cells abolishes the surface exclusion specified by the R plasmid R1drd19, whereas similar disruption in recipients

has no effect, (Beard and Bishop, *J. Bact.*, **123**, 916; 1975), is therefore surprising. Beard and Bishop suggest that surface exclusion is a barrier to exit of DNA from donor cells rather than, as has previously been assumed, to entry of DNA into a recipient. It is presumed that a component (termed ML factor) associated with and dependent on integrity of the mucopolysaccharide in the donor interacts with the *traS* product from the recipient. This leads to prevention of either establishment of stable mating pairs or mobilisation of DNA for transfer. Transport of the *traS* product to the donor must presumably occur

extracellularly by diffusion or by way of the conjugation bridge established during mating. Detailed molecular studies of the cell envelope of male cells should now help to elucidate the precise mechanism of exclusion.

Assuming that a superinfecting plasmid successfully enters a recipient it may then be incompatible with the resident. Plasmids which are incompatible do not necessarily exclude each other. Incompatible plasmids fail to coexist stably in the same cell and one or other is unable to replicate normally, being eventually diluted from the population. Incompatibility also operates in the case of the F factor whether or not the resident plasmid is inserted into the chromosome (Hfr).

Plasmids have been commonly classified on the basis that their incompatibility indicates a close degree of relatedness. Over twenty (in)compatibility groups have been defined (see for example, Chabbert *et al.*, *J. Bact.*, **112**, 666; 1972; Datta and Hedges, *Nature*, **234**, 222; 1971; Grindley *et al.*, *Molec. gen. Genet.*, **119**, 287; 1972). Members of a group are incompatible with each other but almost always compatible with members of other groups. Furthermore, the phylogenetic relationships established by compatibility grouping have been confirmed by studies of base sequence homology on plasmid DNA (see for example, Falkow *et al.*, *J. gen. Microbiol.*, **85**, 65; 1974; Grindley *et al.*, *J. Bact.*, **115**, 387, 1973).

The mechanism of incompatibility is obscure and complex. It may be associated with the process whereby the number of plasmid copies per host chromosome is regulated. Some plasmids, for example the F factor, are present as about one copy per chromosome in *E. coli* whereas others are present as multiple copies. Jacob *et al.* (*Cold Spring Harb. Symp. quant. Biol.*, **28**, 329; 1963) proposed in the replication hypothesis that bacterial chromosomes and plasmids both require specific maintenance sites on the cell membrane for replication and/or segregation. Incompatibility and regulation of plasmid copy number could therefore be explained by competition for a limited number of plasmid-specific sites. This explanation is less satisfactory for multicopy plasmids, which also exhibit incompatibility (Smith *et al.*, *Molec. gen. Genet.*, **129**, 229; 1974) unless either the number of specific sites is high or one master copy of the plasmid responsible for governing replication and segregation binds at a single site. Furthermore, does F incompatibility in Hfr strains mean that the Hfr chromosome occupies the F site or the chromosomal site, or both? To explain such anomalies it has been proposed that plasmid-specific repressors

exist which regulate plasmid copy number. Some support for this comes from studying mutant plasmids which are present as more copies per chromosome than normal (copy mutants). Uhlin and Nordstrom (*J. Bact.*, **124**, 641; 1975) found that copy mutants of the R plasmid R1 fall into two classes. One showed decreased incompatibility with a test plasmid and is assumed to produce a defective repressor of plasmid replication. The other exhibited increased incompatibility due presumably to a defective operator, resulting in increased levels of free repressor. Other mutant plasmids have been isolated which show decreased incompatibility (*Inc⁻*) and allow autonomous replication of superinfecting plasmids (see for example De Vries and Maas, *J. Bact.*, **115**, 213; 1973; Wyman and Novick, *Molec. gen. Genet.*, **135**, 149; 1974). These mutants are however chromosomally inserted. Mutations in chromosomal genes can also affect incompatibility (see for example, Bezanon and Iyer, *J. Bact.*, **123**, 137; 1974).

Hiraga (*Proc. natn. Acad. Sci. U.S.A.*, **73**, 198; 1967) now reports the isolation of novel F' factors that can replicate autonomously in Hfr strains of *E. coli*. These plasmids carry a region of the *E. coli* genome including the *dnaA-bglA* region and a specific site designated *poh⁺* (permissive on Hfr). Since such F' factors effectively bypass the normal *inc*-mediated blocking of plasmid replication, Hiraga has proposed that the *poh⁺* site represents the origin of replication (*oriC*) of the chromosome of *E. coli*. *poh⁺* presumably functions by allowing these F' factors to use *oriC*, instead of the F origin of replication (*oriF*) which is blocked by the *inc* product(s) (repressors?) in Hfr strains. It will be of interest therefore to see whether two autonomous F' factors both carrying *oricC* are compatible. This may determine whether F incompatibility is qualitatively different when the resident F particle is autonomous as opposed to integrated into the host chromosome (Hfr). Complementation studies of mutants of F defective in vegetative replication have been hampered in the past by incompatibility. Therefore the means of circumventing incompatibility afforded by Hiraga's F' factors may aid in the elucidation of the genetic control of plasmid replication.

Masters (*Molec. gen. Genet.*, **143**, 105; 1976) also reports the construction of F' factors carrying the chromosomal region between *bgl* and *mtl* including the origin of replication. It may be significant that this region includes *dnaA*, one of the genes essential for the initiation of chromosome replication. Strains made diploid for the origin of replication by such F' factors grow more slowly than normal and

have a reduced DNA/mass ratio. This is to be expected if the ratio of chromosome origins/cell mass is disturbed since this is thought to be involved in controlling initiation of DNA replication (Donachie, *Nature*, **219**, 1077; 1968). Such strains could be used to determine whether control of initiation of chromosome replication is positive or negative. It should also now be possible to map the precise location of the origin on the genome of *E. coli*. F' factors carrying the origin should therefore prove valuable in understanding the mechanisms of both plasmid incompatibility and the control of DNA replication. □

Ribosomes and secretory proteins

from A. Amar-Costesec

EUKARYOTIC cells contain two distinct populations of cytoplasmic ribosomes: some are free in the cytosol, others are bound to membranes, which thus appear "rough". Cells secreting proteins very actively have a large proportion of their cytoplasmic ribosomes bound to the membranes of the endoplasmic reticulum (ER). In hepatocytes, which synthesise and export serum albumin and other plasma proteins, 60% of the RNA is present in membrane-bound ribosomes and only 20% in free ribosomes. The attached form of ribosomes is predominant in pancreatic acinar cells and in plasmocytes, secreting digestive enzymes and immunoglobulins, respectively. In the mid 1950s, Palade suggested that membrane-bound ribosomes are the site of synthesis of proteins which are to be secreted.

When tissues are homogenised, the ER membranes rupture under the action of shearing forces, but then reform as vesicles. The portions of the ER studded with ribosomes are transformed into rough vesicles studded with ribosomes on their external face. These rough vesicles retain the ability of ER to incorporate amino acids into proteins. Using biochemical and morphological methods, it was demonstrated that ribosomes bind to membranes through the larger of the two ribosomal subunits, the smaller being attached only to the larger (Sabatini *et al.*, *J. molec. Biol.*, **19**, 503; 1966) and that the nascent polypeptide chain is vectorially discharged into the intraluminal compartment of the ER (Redman *et al.*, *J. biol. Chem.*, **241**, 1150; 1966), probably as an extended chain moving through a tunnel in the large ribosomal subunit. The phospholipid membrane forms a barrier between the cytosol and the intraluminal space.

Blobel and Dobberstein (*J. Cell Biol.*, **67**, 835; 1975) propose that ER possesses binding sites in the form of receptor proteins in the membrane. Specific sites on the receptor proteins interact with corresponding sites on the large ribosomal subunit. The receptor proteins form a second tunnel which extends through the thickness of the membrane and is linked to the tunnel of the large subunit. The vectorial discharge of the nascent polypeptide would occur through the two tunnels, resulting in segregation of the protein within the intraluminal space. The main problem, however, is the mechanism by which the cell decides which protein must be segregated, or alternatively how an RNA entering the cytoplasm is able to choose between free or membrane-bound ribosomes.

In 1971 Blobel and Sabatini formulated the 'signal hypothesis' which proposes that mRNAs for secretory proteins possess, after the initiation codon common to all mRNAs, a series of signal codons. When the mRNA is translated by a ribosome, the polypeptide chain then contains at the amino-terminal end a sequence of amino acid residues called the signal sequence. The length of this sequence is estimated at 20–25 amino acids, corresponding to 1,000 to 2,000 daltons. When the nascent polypeptide chain emerges from the large ribosomal subunit, it is recognised by the receptor proteins of the ER membranes, which then form the transient tunnel through the membrane. In the intraluminal compartment, the signal sequence is removed by an endopeptidase. Folding of the peptide chain possibly occurs after this post-translational cleavage. The newly synthesised protein is then ready to follow the pathway of exportable proteins through various cell components (such as Golgi, condensing vacuoles and secretory vesicles).

One implication of the signal hypothesis is that exportable proteins are synthesised as polypeptides larger than necessary, as a result of the presence of the signal sequence. Different authors working with heterologous cell-free systems have observed that the product of *in vitro* translation of mRNA for immunoglobulin light chain is larger by about 20 amino acids than the authentic light chain. Similarly, it has been shown that mRNAs from calf parathyroid glands are translated by a wheat germ extract into parathyroid hormone which migrates more slowly in sodium dodecyl sulphate-polyacrylamide gel than parathyroid or parathyroid hormones. Here also, the translation product is larger than the normal protein. Another implication of the signal hypothesis is that significant homology, if not complete identity, may be expected in the signal

sequence of the polypeptide chains of several exportable proteins. Devillers-Thiery *et al.* (*Proc. natn. Acad. Sci. U.S.A.*, **72**, 5016; 1975) extracted mRNAs from dog pancreas rough microsomes, and translated them *in vitro* in an artificial system. The polypeptides formed were separated by polyacrylamide gel electrophoresis into five "regions". Substantial homology was found in the sequence of 16 amino acid residues at the amino-terminal end. This sequence, which contained predominantly hydrophobic amino acids, could represent the metabolically short-lived signal sequence of dog pancreas secretory proteins. Another compelling argument is given by the work of Schechter (*Proc. natn. Acad. Sci. U.S.A.*, **70**, 2256; 1973) and of Swan *et al.* (*Proc. natn. Acad. Sci. U.S.A.*, **69**, 1967; 1972) who have shown that in mouse myeloma, mRNA coding for the light chain of immunoglobulin is larger than that required to code for a native light chain. It is tempting to believe that the additional nucleotide mass could, at least in part, represent the signal codons.

One advantage of the signal hypothesis is that it resolves the old problem of two functionally distinct populations of ribosomes. All ribosomes are equivalent and whether they become attached to ER membranes or remain free in the cytoplasm depends on the mRNA they are translating. The mechanism proposed for secretory proteins may be applicable to other "segregated" proteins, such as the hydrolases of lysosomes, the oxidases of peroxisomes, and some mitochondrial and chloroplast enzymes. Thus, the signal hypothesis could throw light on the mechanism of biogenesis of complex organelles which contain segregated enzymes. □

Amorphous p-n junctions: curiosity or breakthrough?

from Andrew Holmes-Siedle

THE annoying feature about semiconductor electronics is that we are so frequently forced to use single-crystal forms when it would be ideal if we could instead rapidly deposit thin films of semiconductor on an inert, cheap substrate such as silica glass and perform the usual topological processing on these films (for the many advantages of this method, see, for example, Holmes-Siedle, *Rep. Prog. Phys.*, **37**, 699–816; 1974). Unfortunately, without epitaxial growth conditions, deposited films are always either polycrystalline or amorphous. The grain boundaries

in the former state make it impossible to fabricate many kinds of device while the main bar to the wide use of amorphous semiconductor films has been our inability to form p-n junctions in them by doping.

Doping in semiconductors is the introduction, often in complex patterns, of impurity-containing regions, in order to modify resistivity and conduction type and form rectifying structures. In the crystalline silicon lattice, for example, a phosphorus atom, having a valence bond to spare, contributes electrons to the conduction band. Until recently, it seemed that this effect was always negated in the amorphous network by the large number of free bonds which could accommodate the space valence bonds of a dopant. But Walter Spear and co-workers at the University of Dundee, working with silicon films deposited from a glow discharge in silane gas, found that these films would respond to the presence of dopants, although they were still non-crystalline by all the usual criteria. Now they report the first preparation of a p-n homojunction structure in a completely amorphous material (Spear *et al.*, *Appl. Phys. Lett.*, **28**, 105–107; 1976). This was achieved by beginning the deposition of the silicon film in a stream of silane gas containing a little phosphine and then changing the phosphine to borane, or *vice versa*. The devices which were formed had quite good rectifying characteristics, with 'on-off' current ratios which appear to be about a hundred. The 'on' resistances of fairly large-area devices were still quite large by commercial standards (I would estimate them to be of the order of 10^4 ohms for a square millimetre of junction area) but this might be reduced with further work.

The key to the success of Spear's group probably lies in the careful studies which they previously undertook on the physics of amorphous silicon which included the development of a field-effect measurement of the density of states in the mobility gap against potential energy, thereby making it possible to monitor the variations of this fundamental property while the preparation conditions were varied at the fancy of the investigator. Other aspects of the energy-level spectroscopy of these systems also produced useful guidance as to how nearly the mobility gap and Fermi level corresponded to those in single-crystal material of the same doping. 10^{17} phosphorus atoms give a conductivity of $10 \text{ ohm}^{-1}\text{cm}^{-1}$ in crystalline silicon but only $10^{-4} \text{ ohm}^{-1}\text{cm}^{-1}$ in these films and boron seems to operate at a similar level of inefficiency in doping, implying that the characteristic ability of the amorphous network to negate the doping

Standardisation in human cytogenetics

At the Fourth International Conference on Standardisation in Human Genetics, held in Paris in 1971, a Standing Committee was appointed to examine problems relating to standardisation in human cytogenetics and to make recommendations thereon between international standardisation conferences.

During the past four years the Standing Committee has been responsible for publication of the Report of the Paris Conference and a supplement to that report. In addition, the Standing Committee sponsored a workshop on human cytogenetic registries, held in Edinburgh in April 1975 (*Nature*, **256**, 456; 1975) and on the recommendation of this workshop, appointed an international advisory committee on cytogenetic registries.

The terms of reference of the Standing Committee call for elections to a new committee no later than the International Congress of Human Genetics in October, 1976 and the possibility of holding the next standardisation conference at that time. For these reasons the committee has recently circulated recommendations to the participants of the Paris Conference, which have been approved as follows: (1) That a standardisation conference on the lines of the Paris and Chicago Conferences could not be justified in 1976 and, therefore, that no such confer-

ence should be held in conjunction with the Fifth International Congress of Human Genetics in Mexico City in October, 1976; (2) That a Standing Committee charged with the task of monitoring developments relating to standardisation and nomenclature in human cytogenetics should be maintained; (3) That an open meeting of all interested human cytogeneticists should be held at the Fifth International Congress of Human Genetics in Mexico City in October, 1976. At this meeting a report of the Standing Committee would be received and a new Standing Committee elected. The existing Standing Committee will be acting as a nominating committee and will be preparing a slate of nominations for consideration by the meeting in Mexico City. These nominations will take into account regional and sectional interests to ensure, so far as possible, a fully representative Standing Committee.

Suggestions for nominations should be sent to the Standing Committee. The time and place of the Mexico City meeting will be announced later and circulated to those requesting this information. Nominations and requests for further information should be addressed to: Dr John L. Hamerton, Department of Genetics, Health Sciences Children's Centre, 685 Bannatyne Avenue, Winnipeg, Manitoba, R3E 0W1 Canada.

effect is still prevalent, since roughly 999 out of 1,000 dopant atoms still seem to be suppressed. It remains to be seen whether this apparent inefficiency in doping can be improved on. The important discovery, however, is the fact that even a few dopant atoms are not suppressed. Junction action is thereby produced and a new device technology may even be in the making. The extent to which this technology could compete with single-crystal semiconductor technology depends less on the inefficiency of doping mentioned above, which can be countered, than on the achievement of low minority-carrier recombination-generation rates in the junction regions. High rates lead to high reverse currents in rectifiers and loss of efficiency in photovoltaic applications. It will be a much more challenging task to make these recombination rates suitably low but success in doing so could have a greater impact than the much-researched 'amorphous threshold switch' technology, which now seems to have lost any commercial momentum

it ever had. An amorphous p-n junction technology would have much in common with the increasingly popular silicon-on-sapphire technology except that the serious restrictions in area imposed by the dimensions of sapphire crystals would have been removed and the way thus opened for integrated circuits as large as the page on which this column is written. □

Superfluid within superfluid

from P. V. E. McClintock

At last, somebody has gone out on a limb and named an encouraging number for the temperature of the anticipated superfluid transition among the ^3He atoms of a liquid ^3He - ^4He solution. According to Patton and Zaringhiam of the Massachusetts Institute of Technology, writing in a recent issue of *Physics Letters* (**55A**, 95; 1975), the transition is likely to

take place under conditions which can be achieved experimentally.

They are brave men. The first estimates of the superfluid transition temperature T_c of pure liquid ^3He were wrong (too high) by a factor of more than fifty. Experimenters eagerly sought the phenomenon, but failed to find it at the predicted temperature: the theorists then increased the sophistry of their calculations and revised their estimates of T_c downwards, a cycle which was repeated several times. This procedure was so utterly demoralising for those doing the experiments (being the classical brainwashing technique of repeatedly raising the victim's hopes and then dashing them again) that, when the superfluid transition finally did manifest itself in Osheroff's cryostat at Cornell, it took several weeks for him and his co-workers to realise the significance of what they had discovered. Hence the notable reluctance, so far, of most theorists to predict a T_c for ^3He - ^4He solutions.

A solution of a few percent ^3He in ^4He is stable down to the lowest temperatures and is, even without the expected transition, of considerable intrinsic interest. The ^3He atoms exist and move in a background aether—the liquid ^4He —which is itself a superfluid whose transition temperature lies at the relatively high value of 2 K. At low temperatures, where the ^4He entropy is essentially zero, the ^3He atoms dominate the behaviour of the liquid. In fact, if one completely ignores the presence of the ^4He (except insofar as it results in each ^3He atom having an effective mass which is larger than the bare atomic mass), and simply treats the ^3He atoms as though they were an ideal gas, one gets the right answers for most of the liquid's properties.

Now, it is known that an attractive interaction occurs between the ^3He atoms of a solution and, at a sufficiently low temperature, this is expected to give rise to the formation of Cooper pairs, and hence to superfluidity. It is not yet clear whether the relative angular momentum of the atoms in a pair will be zero, as for the Cooper pairs of electrons in a superconductor, or unity as is apparently the case in superfluid pure ^3He . The situation would certainly be very peculiar in either case, however, since the superfluid ^3He which was formed would exist in another, completely separate but interpenetrating superfluid, the ^4He , which is fundamentally different in nature in that it does not acquire its superfluidity through the formation of Cooper pairs. The bizarre physical properties to be expected of such a system were discussed in considerable detail by Khalatnikov in his invited address (Khalatnikov, Mineev and Volovick:

Proc. 14th. Inf. Conf. on Low Temp. Phys., North Holland Pub. Co., Amsterdam, **5**, 102; 1975) to the recent LT14 conference in Helsinki—but he was careful to avoid alluding to any numbers. In particular, he did not make any mention of a value for T_c .

Patton and Zaringhalam have now described a theory which, with one adjustable parameter, gives a reasonable description of the pressure dependence of T_c for pure ^3He . They fix this parameter by fitting their theory to the experimental results already obtained and then, together with a variety of other experimental information, use it to derive a value of T_c for a ^3He - ^4He solution. They conclude that, for a saturated solution of 8.85% ^3He in ^4He under a pressure of 20 bar, $T_c = 0.9$ mK. This result is some 10,000 times larger than an earlier prediction by Østgaard (*Phys. Lett.*, **94A**, 433; 1974) of $\sim 10^{-4}$ mK.

These calculated values of T_c may be compared with 0.7 mK, the lowest temperature to which pure ^3He has so far been cooled. Making thermal contact to a ^3He - ^4He solution is rather more difficult than to pure ^3He at these temperatures, however, so that cooling a solution below 0.9 mK may not necessarily be quite within present experimental capabilities; but it certainly cannot be far removed.

Remembering the bitter experience of the theoretical T_c values for pure ^3He , a certain level of scepticism is inevitable in relation to Patton and Zaringhalam's 0.9 mK. It is nonetheless encouraging, and we should not have to wait very long for their prediction to be put to the test in an actual experiment. □

No need for glyophorin?

from a Correspondent

It is a truism that any breakthrough in science, as in art, fashion or sport, quickly becomes a commonplace. Nowhere is this more obvious than in protein sequencing. However, although the complete sequencing of a protein no longer has, perhaps quite the same impact as previously, the sequencing of a membrane protein is still a relatively rare occurrence. This has been achieved recently for glyophorin, a major protein of the human erythrocyte membrane (Tomita and Marchesi, *Proc. natn. Acad. Sci. U.S.A.*, **72**, 2964–2968; 1975; see *News and Views*, **258**, 478; 1975). Some of the sequence is familiar from earlier publications by Marchesi and his colleagues. In par-

ticular the now famous sequence of 23 amino acids lacking charged residues and enriched in hydrophobic residues such as leucine, isoleucine and valine, is placed firmly towards the carboxyl terminal of the polypeptide chain. This sequence is thought to be responsible for integrating the protein into the hydrophobic environment of the membrane. A segment of the polypeptide chain at the carboxyl end is believed to protrude internally whereas the amino terminal sequence, bearing a large amount of carbohydrate, is on the outer side of the membrane.

The function of this quantitatively important constituent of the erythrocyte membrane is unknown. Some recent results suggest that it may be dispensed with altogether without undue effect on the survival either of the erythrocyte or of the person carrying these cells.

The starting point for this work was the identification of a rare genetic variant in humans which affects erythrocyte surface antigens. The En^a antigen is present on the surface of almost all human erythrocytes except in some very rare individuals. Three families are known with this trait: two in Finland and one in England (Dainborough *et al.*, *Vox Sang.*, **17**, 241–255; 1969; Furuholm, *et al.*, *Vox Sang.*, **24**, 545–549; 1973). These individuals, who are clinically normal, also show unusually weak activity of another common erythrocyte antigenic system called MN. It has been known for many years that MN antigenicity is associated with glyophorin and this prompted Tanner and Anstee (*Biochem. J.*, **153**, 271–277; 1976) to look for this component in En(a–) erythrocytes. It turns out that the erythrocytes apparently lack glyophorin completely. Erythrocyte En(a–) membranes analysed by SDS polyacrylamide gel electrophoresis have normal peptide composition except for the loss of glyophorin and a minor glycoprotein species of slightly smaller molecular weight.

Tanner and Anstee then tried to label glyophorin in En(a–) erythrocytes with lactoperoxidase. In normal erythrocytes, glyophorin and another membrane protein are readily iodinated by this means. In En(a–) erythrocytes however, only one membrane component was labelled and this was not glyophorin.

What these results imply, therefore, is that glyophorin itself is not absolutely required for the function of the erythrocyte. This is surprising on two counts; first, glyophorin is one of the two most abundant proteins of the erythrocyte membrane. Together with a component called protein 3 it accounts for almost half of the membrane proteins. Second, there is considerable evidence that glyophorin and protein

3 form a non-covalent complex in the membrane. It has been suggested (see Rothstein *et al.*, *Fedn Proc.*, **35**, 3–10; 1976) that the complex plays a coordinated role in anion transport. Although protein 3 appears to be the more directly involved in anion transport, it will be interesting to see if the quite drastic changes in the composition of membrane protein of En(a–) erythrocytes result in any physiological effects on ion transport. □

Aspects of tumour biology

from M. J. Bevan and I. S. Trowbridge

The Armand Hammer Symposium in Tumour Biology was held at The Salk Institute, San Diego in January 1976.

THE origin of the cancer cell and its interaction with the immune system were the main themes of the Symposium.

In the opening session, David Baltimore (Massachusetts Institute of Technology) and Charles Heidelberger (McArdle Laboratory, Madison) discussed viral and chemical oncogenesis respectively. Baltimore showed that the resistance to infection by N-tropic or B-tropic murine leukaemia viruses conferred by the FV-1 locus is probably at the level of transcription implying that FV-1 is a regulatory gene coding for a repressor. Viral DNA could be made and integrated into the genome of resistant host cells but unlike the case of sensitive host cells virtually no viral RNA was produced. He also described the *in vitro* transformation of murine bone marrow and foetal liver cells by Abelson Leukaemia virus and its quantitation by a focal assay. Whether the target cell for transformation is a pre-B cell or a more primitive stem cell and how many such cells are present in the various haemopoietic organs is uncertain. It seemed unlikely that a B-cell itself is the target as no immunoglobulin-bearing cells were transformed.

Heidelberger established that chemical carcinogens actually transform normal cells rather than select pre-existing malignant cells from within the population of "normal" cells in his *in vitro* fibroblast cultures, by showing that frequencies of transformation greater than 50% are possible. He also described evidence that chemical carcinogens do not cause transformation

by activating endogenous RNA tumour viruses. One importance of his experimental system is that it provides a screening procedure for potential carcinogens. In this regard it was interesting to hear that the effects of promoters, chemicals that do not themselves cause cancer but which potentiate the action of other carcinogens, can also be assayed in this system.

George Klein (Karolinska Institute, Stockholm) described the interaction of Epstein-Barr virus with human B lymphocytes. He described convincing evidence for the identity of EB virus receptor and the complement receptor of B cells stressing however that not all complement receptors bind the virus. He also showed experiments with EB virus-negative B cell lines and their *in vitro* transformed sublines which illustrated the selective *in vitro* growth advantage that transformation with EB virus confers on the transformed cell.

Gerald Edelman (Rockefeller University, New York), suggested that it may be possible to bring together some of the diverse changes which take place when a cell is transformed by considering the role submembranous microtubule assemblies have in maintaining cell morphology and transmitting extracellular growth signals to the cell interior.

Many tumours express cell surface structures which can be recognised as foreign by the immune system of the host. Ways in which genes in the major histocompatibility complex may control immune responsiveness were a major theme. J. Klein (University of Texas, Dallas) showed partial amino acid sequence data from four laboratories on various alleles of the 45,000 molecular weight H-2K and H-2D major histocompatibility antigens of the mouse. Comparison of the sequences of different alleles of one locus shows multiple amino acid replacements—30–40% of the positions are different. These are “complex allotypes” in Lee Hood’s terminology though Klein warned that comparing products of other alleles with the H-2^b antigens could be misleading since this haplotype may have come from a different subspecies or different species of mouse than that from which other inbred strains were derived. Comparison of the sequences of the products of the K and D loci again shows 20–40% amino acid substitutions. Here the remarkable thing is the degree of similarity between their products, confirming their origin by gene duplication. The K and D sequences of the mouse appear more similar to each other than to either the human HLA-A or HLA-B sequences, that is, species-specific residues are shared by K and

D. This raises the possibility that the two loci, K and D in mice and HLA-A and HLA-B in humans separated in evolution after divergence of the species.

What does this have to do with immune responsiveness? These gene products control interactions between cytotoxic effector T cells and target cells. Doherty and Zinkernagel have suggested that neoantigens (including tumour-associated antigens) are recognised only in association with K and D products as complex determinants (the altered-self hypothesis). Klein favoured the idea that the K and D restriction of cytotoxicity is explained by two independent interactions. First, the T cell receptor binds the neoantigen which only leads to a successful cytolytic interaction when the associated H-2K or D itself binds the identical K or D molecules on the target cell, a like-like interaction. Edelman felt that this was ruled out by his experiments with Schrader showing that antibody to H-2K and D blocked a cytolytic interaction only when the target (not the killer cell) was treated. He proposed a T-cell receptor with two subsites, one for K or D and the other for neoantigen leaving the origin of the restriction as “altered self”.

H. McDevitt (Stanford University, California) described the distribution of H-2 I-region-associated antigens on various T cell populations. Neither precursor cytotoxic T cells, effector cytotoxic T cells, nor helper T cells are removed by treatment with anti-Ia serum and complement. In contrast, suppressor T cell activity in two systems (allotype suppression and suppression of a hapten carrier response) is sensitive to anti-Ia serum and complement. Partly on the basis of these results, McDevitt interprets the helper T cell replacing factor of Taussig and Munro, which is Ia positive and antigen specific, as being a suppressor T cell factor like that described by Tada. The factor, according to McDevitt, enhances the IgM response only as a consequence of suppressing the IgG response.

Despite this, McDevitt believes that Ir genes which determine high or low responsiveness to particular antigens, are expressed at the helper T cell level, that is, failure to develop helper T cell activity is not due in all cases to excessive suppressor T cell activity. Therefore the Ir gene product expressed in helper T cells is predicted to lack Ia determinants.

A. Mitchison (University College London) was concerned with whether two immunogenic but separate molecules on the same cell would be co-processed by the immune system. Some tumour antigens may be weak because the tumour cell presents no

other determinants for associative recognition. He showed that the antibody response to the Thy-1.2 alloantigen could be enhanced by the associative recognition of other H-2 coded or non-H-2 coded alloantigens. Nylon wool purified primed T cells do help the anti-Thy-1.2 response of primed purified B cells and Mitchison has evidence that the T cells and B cells effectively cooperate when recognising different alloantigens on the same immunogenic cell. AKR (H-2^k, Thy-1.1) spleen cells which have been primed with CBA thymocytes (H-2^k, Thy-1.2) give a good response to CBA thymocytes on transfer into irradiated AKR recipients. B10.BR thymocytes (H-2^k, Thy-1.2) also stimulate a good secondary response but B10 (H-2^b, Thy-1.2) or F₁ (CBA × B10) thymocytes fail to generate a good response. This failure to stimulate cannot be interpreted as meaning that Thy-1 is seen by T cells as a modification of H-2 antigens, that is, as a type of altered-self hypothesis. Mitchison views this as an example of antigenic competition in which recognition of a foreign H-2 somehow removes the Thy-1 antigen (which has to be on the same cell) from the inductive processing system.

Antigen presentation in the response of rats to syngenic chemically induced tumours was a feature of R. Baldwin’s talk (University of Nottingham, UK). An animal protected against a tumour by immunisation with irradiated cells shows responsiveness to individually specific tumour antigens. A tumour bearer, on the other hand, responds also to embryonic antigens common to many tumours (as detected in the lymph node microcytotoxicity test). Grafting irradiated tumour cells leads to tumour-specific protective immunity. However, if the host has been pre-exposed to tumour antigen in the form of isolated membrane or 3M KCl extracts then a subsequent graft of irradiated tumour cells does not lead to protective immunity although a humoral response is observed. This raises questions of how the class of the response to a tumour is determined.

Melvin Cohn (The Salk Institute, San Diego) made the interesting comment that all tumours presently used as model systems have already undergone rigorous selection for their ability to escape immunological rejection since they arise in immunocompetent hosts. He suggested that the complexity of tumour rejection may be more easily analysed by establishing tumours in immunoincompetent hosts from the *in vitro* transformed colonies of Heidelberger and studying the way in which these cells escape single step immunological selection pressures such as T cell killing or antibody-dependent cytotoxicity. □

review article

Archaean high grade complexes and modern continental margins

B. F. Windley* & J. V. Smith†

Two kinds of geological terrain formed before 2,500 Myr ago. First, the greenstone belts, which are linear basins full of sediments and volcanics, possibly comparable with the marginal basins that occur behind modern island arcs. Second, and the subject of this paper, are high grade complexes containing rocks that have been deformed and recrystallised at high temperatures and pressures—granitic gneisses derived from granitic rocks are particularly common. Examples occur in north-western Scotland, Greenland, the Labrador coast, the Limpopo Belt in southern Africa, southern India, and probably the Aldan and Kola shields in the USSR. These complexes are so similar lithologically that broad generalisations about their tectonic evolution are justifiable. No major attempt has been made to see if they share any similarities with modern geological environments whose origins are better understood. We suggest that they are ancient analogues of the modern deep seated granitic batholiths located along the main arc axis of active plate margins of North and South America, for example in California and the Andes; accordingly, they are complementary to the greenstone belts in their back-arc setting.

THE tectonic evolution of Archaean high grade gneissic regions is poorly understood, especially the origin of tonalitic–granodioritic gneisses which make up 80–85% of the present surface^{1,2}. Various source materials have been proposed, on the basis of chemical analogies; they include andesitic–dacitic volcanic rocks^{3–6}, modified earlier gneisses⁷, dioritic rocks that underplated the crust⁸, plutonic granitic rocks in general^{1,9,10}, and igneous tonalites and granodiorites¹¹. Some local chronologies are known^{9,10,14}, but little is understood of their evolutionary meaning.

We propose a more general approach in which all the tectonic and chemical interrelations of the Archaean rock types are compared with properties of rocks in Mesozoic and Cainozoic fold belts, whose tectonic environments are better understood. Specifically, we suggest that Archaean quartzofeldspathic gneisses and their enclosed remnants of layered igneous complexes¹² and supracrustal rocks are metamorphosed and deformed equivalents of calc-alkaline batholiths in the Cordilleran Fold Belt of western North America, and probably of the similarly active plate margin of the Caledonian Fold Belt of Britain. A strict uniformitarian comparison cannot be expected because, first, the older rocks show indications of metasomatism; second, mineralogical reactions have occurred which indicate burial of some of the Archaean rocks to 20–30 km; third, earlier plates may be smaller, less coherent and more mobile than modern plates, resulting in a different tectonic style at continental margins; and fourth, early magmas produced by partial melting of the lower crust and upper mantle may differ chemically from modern ones. Even at the risk of over-

simplification, however, we have made the comparisons in order to provide ideas to be tested by further mineralogical, petrological and geochemical studies.

Archaean relationships

The Archaean high grade regions, to which this model is most readily applicable, are north-western Scotland (the Scourian)¹³, Greenland^{1,10}, Labrador¹⁴, the Limpopo Belt of southern Africa¹⁵, and southern India¹⁶.

Gneisses formed at least twice in Greenland and Labrador, and only the younger are relevant here. In the other regions only the younger gneisses have been recognised so far. The relevant chronology is ^{1,9,10,14}:

Oldest	Formation of gneissic basement
	Deposition of supracrustal rocks (basic volcanics, pelitic schists and gneisses, marbles and quartzites)
	Interthrusting of supracrustals and basement
	Intrusion of layered ultramafic–gabbroic–anorthositic complexes
	Intrusion of tonalite–granodiorite
Youngest	Deformation and metamorphism of tonalite–granodiorite to give rise to high grade gneisses.

The Greenland gneisses grade into homogeneous ‘granitic’ rocks in low-deformation zones⁹, and Labrador gneisses occur as deformed veins of intrusive ‘granite’¹⁴. Gneisses in all regions have calc-alkaline affinities whether metamorphosed to amphibolite or granulite grade. Accepting the arguments against a volcanic origin⁸, we prefer to envisage a suite of tonalitic–granodioritic rocks as the precursor of the quartzofeldspathic orthogneisses¹. The chemistry of major and trace elements in the gneisses compares best with that of ‘granitic’ rocks from

Addresses: *Department of Geology, University of Leicester, Leicester, UK.

†Department of Geophysical Sciences, University of Chicago, Chicago, Illinois 60637.

modern continental margins¹¹, and the very low initial ratios of strontium isotopes (about 0.701) are consistent with direct derivation from the mantle¹⁷.

Metamorphosed remnants of layered igneous complexes with an idealised upward stratigraphy of ultramafic–gabbro–anorthosite¹² occur within the gneisses. The complexes are bordered either by supracrustal rocks into which they were intruded¹⁸ or directly by the gneisses¹⁹. Parts of the Fiskenaeset Complex, western Greenland, are associated with supracrustal rocks²⁰, and other parts are not¹⁹. The intrusion of the complexes was followed immediately by that of the ‘granitic’ rocks^{9,10,14}, resulting in the complexes occurring as giant rafts in vast quantities of ‘granitic’ rocks: thus all the Fiskenaeset gneisses, accounting for 84% of the present surface area², are younger than the Fiskenaeset layered complex²¹. Even where some older gneisses predate the complexes, as near Godthaab, western Greenland, the younger gneisses may account for over half the area⁹. Apparently the most highly developed gneisses coincide with the most highly developed layered complexes (for example, at Fiskenaeset).

The complexes are in all stages of fragmentation. In places, individual layers persist for many kilometres (for example, Fiskenaeset, Limpopo), but the complexes commonly occur as isolated lenses and pods from kilometre to metre size (southern India²², western Greenland²³, Scourian¹³). Although late deformation was responsible for local folding, thinning and boudinage of some layers, it would seem an unlikely cause of the fragmentation of the complexes into the millions of small pods and lenses found in many areas. Perhaps this fragmentation was connected with the engulfment of the complexes by granitic material on an enormous scale. Explosive mechanisms may have been involved, and a detailed comparison of the Archaean occurrences with the appinitic rocks and breccias of the Caledonian mountain chain, especially in Donegal, is desirable¹⁰.

Comparison with oceanic crust

Archaean sediments, volcanics and layered complexes have been interpreted as representing a section through oceanic crust²⁴. Graphitic pelites and marbles have been correlated with modern pelagic sediments leading to the interpretation of quartzites and psammites as recrystallised oceanic cherts; the occurrence of dolomitic limestones at the Mid-Atlantic Ridge has also been noted. Amphibolites with occasional relict pillows have been matched with oceanic basic volcanics, and the anorthosite–leucogabbro–ultramafic sequence has been related to layered igneous complexes of the type inferred to occur under some modern oceanic basalts²⁵.

Objections can, however, be raised. Where the Archaean quartzites are best developed, as in the Limpopo Belt, they are current-bedded and thus interpreted as orthoquartzites (R. M. Shackleton, personal communication). The carbonate (dolomitic marble)–orthoquartzite–K-rich pelite (sillimanite mica schist) association, typical of most high grade regions, is best explained by deposition in a shallow-water, continental margin^{26,27}. Comparison of the chemistry of amphibolites with that of volcanic rocks is fraught with difficulties because of chemical changes undergone during metamorphism²⁸. Some Archaean amphibolites have chemical patterns similar to those of modern island-arc volcanics (B. E. Leake and B. F. W., unpublished) and others to oceanic volcanics²⁹. To distinguish the latter from the early tholeiitic series of island arcs is difficult, especially after metamorphism³⁰. Anyway, low-K tholeiites occur in various tectonic environments, and distinction between oceanic and island-arc or continental volcanic rocks, so demonstrable in modern times, should be less apparent in the Archaean³¹. Grab samples indicate that layered igneous complexes in modern oceanic crust apparently comprise predominantly Ti, Fe-rich gabbro, and contain only about 5% Cr-poor labradorite anorthosite and gabbroic anorthosite (leucogabbro)²⁵. Archaean complexes are very different in both rock proportions and chemistry; thus gabbros are minor and not enriched in Ti and

Fe, whereas the bulk of the bodies consist of bytownite–anorthosite and leucogabbro (where best preserved), often inter-layered with chromitite. Alpine-type ophiolite complexes derived from the oceanic crust are distinguished by a scarcity of the cumulate features characteristic of gravity settling, whereas many Archaean complexes contain chromitites, plagioclase cumulates and crystal-graded layers. Volcanic and plutonic rocks formed at mid-ocean ridges should be characterised by highly variable age relations, stratigraphic succession and distribution (J. Cann, personal communication), whereas Archaean rocks, especially in western Greenland and Limpopo, are noted for remarkable stratigraphic regularity and persistence along strike for tens of kilometres. Only a little acidic tonalite can be generated by one-stage fractionation of ultrabasic mantle.

We believe that these differences between recent oceanic rocks and Archaean complexes are so great that the latter cannot be slices of oceanic crust.

Comparison with Cordilleran fold belts

The Mesozoic batholiths along the axis of the Cordilleran fold belts of North and South America represent the roots of magmatic arcs related to palaeosubduction zones. The most abundant rocks are tonalite, granodiorite and quartz monzonite (adamellite)³². The content of K₂O generally increases away from the Pacific margin³³ although potassic granites are a minor component of the batholiths³⁴.

Within the Peninsular Range batholith of southern California, layered igneous complexes composed largely of calcic plagioclase and primary amphibole include a succession of minor peridotite and troctolite and predominant hornblende anorthositic gabbro³⁵. The mineralogy and chemistry of these rocks is similar to that of the Archaean complexes. The California complexes have been interpreted as early cumulates of a hydrous and refractory nature precipitated from a basaltic–andesitic magma which later yielded the calc-alkaline batholithic suite.

Concentrically-zoned ultramafic complexes of the Alaskan type are clearly distinguishable from the alpine-type ultramafic complexes usually considered to be obducted oceanic crust and mantle. The Alaskan-type complexes, which also occur in the Ural Mountains, British Columbia, Venezuela and perhaps Japan, have been interpreted as the products of fractional crystallisation and flow differentiation of basaltic magma in deep-seated, volcano feeder pipes³⁶. Crystallisation under hydrous conditions in the pressure range 2–11 kbar has been invoked to explain the presence of magnetite and hornblende as well as olivine, clinopyroxene and plagioclase. The associated eruptive products are andesites and probably aluminium-rich tholeiitic basalts.

Although few detailed mineralogic–petrologic studies have been undertaken, it is certain that calc-alkaline volcanic and plutonic rocks are linked with the hydrous, amphibole-bearing intrusions, typically developed along modern continental margins.

Many of the layered complexes within the Peninsular Range tonalitic batholith are bordered by a group of sediments (the Julian Schists) that comprise sillimanite mica schists (meta-K pelites), quartzites and carbonate rocks, an assemblage that is strikingly similar to that in the Archaean tonalitic gneisses and that likewise formed in a shelf–continental margin environment.

In the Dawros–Cashel area of Connemara, several fragmented layered complexes³⁷ with peridotite^{38,39} and hornblende–calcic plagioclase gabbro grade upwards into plagioclase gneiss⁴⁰. According to a plate tectonic model⁴¹, this area lies on the northern continental margin of the Caledonian Fold Belt. We suggest that the petrochemical features of these bodies render them comparable to the Archaean and California examples. Particularly important is the generally accepted evidence of major horizontal movements—now represented by the extensive occurrence of nappes—together with the regional metamorphism to garnet and kyanite grades in the Dalradian rocks of Scotland⁴². These features occur in the high grade Archaean

complexes¹⁰, and apparently represent the final stage of tectonic activity at an active continental margin.

Mineralogical–petrological–chemical considerations

The mineralogy and petrology of the Archaean complexes is interpreted as the result of igneous crystal–liquid differentiation combined with deformation and regional metamorphism. The Fiskenaesset Complex provides the clearest evidence. Correlation of folded fragments has revealed seven major zones. From bottom to top they are: minor, magnetite-rich ultramafic rocks; leucogabbro with anorthosite laminae; gabbro; leucogabbro with marked cumulate texture; anorthosite; chromitite; and garnet anorthosite^{20,43}. Hornblende occurs in all of the rock types. This sequence has been interpreted as the result of gravity differentiation of a water-rich basaltic magma in which the early crystallisation of Mg–Al–Ti–Cr–magnetite and the continued precipitation of primary magnesio- to tschermakitic-hornblende led to a trend quite different from that found in dry, layered basaltic complexes such as Skaergaard. Some megacrysts may locally represent original primocrysts, but most rocks have metamorphic textures superimposed on modal variations retained from the original igneous layering. Related structural and mineralogical evidence has been found in the Sittampundi (ref. 44 and B. F. W. and J. V. S., unpublished), Limpopo¹⁸ and Lewisian complexes⁴⁵, but the data are less complete for various reasons. Solid-state recrystallisation has modified mineral compositions in all of the complexes, and metasomatism has changed some bulk compositions.

Although estimates of the temperatures and pressures reached during metamorphism depend on somewhat controversial thermodynamic models and experimental phase equilibria, metamorphism of the Archaean high grade complexes certainly occurred at depths estimated at between 15 and 40 km. In South Harris, chemical relationships between garnet, pyroxenes and olivine indicate that granulite metamorphism occurred near 800–860 °C and 10–13 kbar (ref. 46). Garnet-clinopyroxene rocks from Sittampundi reacted near 850 °C and 7–8 kbar (ref. 47). Also at Sittampundi, the coexistence of staurolite, kyanite, garnet and other minerals in one hand specimen indicates conditions near 600 °C and over 5 kbar, whereas the occurrence of sillimanite-bearing assemblages indicates somewhat different conditions (B. F. W. and J. V. S., unpublished). Olivine and plagioclase grains have not been found together in a single hand specimen from Fiskenaesset, and they may have been replaced by spinel, pyroxene and amphibole. Adjacent grains of orthopyroxene and plagioclase coexist in some Fiskenaesset rocks without having reacted to garnet and quartz. Chemical relationships between finely intergrown oxide minerals (B. F. W., J. V. S., I. M. Steele and F. C. Bishop, unpublished) indicate equilibration down to 300–600 °C, and those between intergrown sulphide minerals (F. C. Bishop, J. V. S. and B. F. W., unpublished) indicate equilibration down to 100–200 °C. The metamorphic history differs from one complex to another, from one part to another, from one mineral assemblage to another, and from one grain to another. In spite of the local idiosyncrasies, which depend on relative diffusion coefficients and stresses, there are overall large scale effects interpretable in terms of high grade regional metamorphism which must have affected the tonalite–granodiorite gneisses as well as the igneous complexes.

The extent of metasomatism is uncertain. In the Fiskenaesset Complex, mineral compositions vary monotonically with position⁴³, and erratic metasomatism should be insignificant except perhaps for minor development of mica and sodic plagioclase and of some minerals of the contact facies²⁰. In the Limpopo Complex, one traverse showed strong erratic plagioclase zoning to more sodic compositions, whereas compositions were steady in another traverse¹⁸. A major question is whether most of the amphibole in the Fiskenaesset Complex crystallised from the igneous magma or whether it is a metasomatic product

after pyroxene. In some thin sections, amphibole occurs around pyroxene grains, but the intimate interlayering of amphibole-free and amphibole-bearing assemblages in many felsic and mafic rocks suggests an igneous origin before metamorphic recrystallisation. Although small composition differences between amphiboles from felsic and mafic parts of the same zone⁴³ probably result from metamorphic equilibration, the generally tschermakitic-pargasitic nature of all the amphiboles is like that of amphiboles produced synthetically from hydrous basaltic melts (see, for example, refs 48–51). Furthermore, tschermakitic-pargasitic hornblendes occur in modern unmetamorphosed rocks where the texture is consistent with the crystallisation of amphibole from a magma (see refs 52–54). Rare symplectites of spinel and pyroxene may result from the dehydration of amphibole²⁰. Although much further study is needed, especially of relatively undeformed specimens from low deformation zones, we assume here that amphibole is one of the igneous minerals at Fiskenaesset, and that the magma was hydrous.

Magnetite occurs as dust in some mafic silicates at Fiskenaesset, but most magnetite of the lowest zone occurs as discrete Cr-rich grains whose association with ilmenite and Mg–Al spinel can be interpreted as the result of metamorphic decomposition of primary, Mg–Al–Cr–Ti-bearing magnetite (I. M. Steele, F. C. Bishop, B. F. W. and J. V. S., unpublished). Its occurrence throughout the complex as a primary phase, especially in the early mafic zone, would explain why there is no strong trend towards iron enrichment. The primary precipitation of magnetite requires redox conditions more oxidising than those within dry basaltic layered complexes⁵⁵, and redox conditions for the Fiskenaesset magma could be inferred by reference to phase relationships in basaltic melts⁵⁶. Chromium-bearing spinel, loosely called chromite, occurs in many Archaean complexes. Its composition range—atomic Mg/(Mg + Fe) 0–0.4; Cr/(Cr + Al) 0.2–0.6—distinguishes it from Cr-spinels in alpine-type, concentric-Alaskan, and dry-basaltic layered complexes (ref. 18, and J. V. S. *et al.*, unpublished). The compositional adjustment of cumulus Cr-rich spinels by reaction with silicates and trapped liquid has been demonstrated in the Rhum layered intrusion⁵⁷. No zoning was detected in any Fiskenaesset chromites in spite of considerable textural variation, and the general compositional features are independent of the modal proportion of Cr-spinel to silicates. Particularly noticeable is the absence of Cr-spinel in most, but not all, of the lowest ultramafic zone at Fiskenaesset, and its principal occurrence as thin layers between late anorthosites. Many basaltic intrusions have early chromite and late magnetite. The near-absence of early chromite at Fiskenaesset can be explained by the precipitation of Mg–Al–Cr–Ti-magnetite as a result of a relatively high state of oxidation.

Archaean high grade complexes carry highly calcic plagioclase (Fiskenaesset, mostly An_{96–80} (ref. 43); Limpopo, up to An_{75–80} (ref. 18); Sittampundi, up to An₁₀₀ (ref. 44, and J. V. S. *et al.*, unpublished)) which distinguishes them from Proterozoic anorthositic complexes carrying labradorite and andesine. Even the latest plagioclase at Fiskenaesset is highly calcic (An⁸⁰), and there is no trend to very sodic or even potassic feldspars, as in many differentiated basalts. As proposed for the ultrabasic and basic intrusions of Connemara³⁷, the most obvious explanation for the persistently high Ca content of Fiskenaesset plagioclase is the continued depletion of Na and K from the wet residual magma by precipitation of primary hornblende with ~2% Na₂O, ~0.4% K₂O and ~12% CaO. Calcium-rich pyroxenes from dry basaltic complexes carry <1% Na₂O, <0.2% K₂O and ~18% CaO, whereas Ca-poor pyroxenes carry hardly any Na₂O and K₂O, thereby permitting the concentration of Na and K into later plagioclase.

In spite of all the uncertainties caused by regional metamorphism and deformation, the major mineralogical features of the Fiskenaesset complex seem to be qualitatively consistent with the crystal–liquid differentiation of a wet, oxidised, basaltic magma in a closed magma chamber. Detailed experimental

syntheses along the lines of those already carried out on basaltic and ultrabasic compositions (refs 48–50, and 58–61), followed by subsolidus equilibration, are needed to provide detailed reference data. In the meantime, it is necessary to rely on qualitative and semiquantitative features (the early precipitation of Mg–Al–Ti–Cr-magnetite; late crystallisation of highly calcic plagioclase and high-Fe, medium-Cr spinel; prolonged crystallisation of tschermakitic–magnesian–hornblende; medium Fe/Mg ratio of ferromagnesian minerals) in looking for possible matches to the Fiskenaesset Complex in modern environments. To set the search in the widest context, modern rocks with highly calcic plagioclase need to be examined, and their wider mineral contents should be checked against the Fiskenaesset Complex.

A nodule from alkali-olivine basalt, Auckland, New Zealand, contains 74% An_{97} (ref. 62). Ultramafic nodules in basalts from Hawaii and Itinome-gata, Japan, carry minor plagioclase ranging up to An_{90} (refs 63 and 64). Associated minerals are too highly magnesian to match those at Fiskenaesset. Some basalts from oceanic environments carry megacrysts or phenocrysts of composition An_{70-89} (refs 65–67), but the differentiation trend in these near-surface flows is quite unlike that in Archaean complexes. The younger volcanic rocks of Tonga, and the Cape Hoskins volcanic rocks, Papua, contain An_{80-90} phenocrysts^{68,69}, but titanomagnetite occurs only in the late differentiates; amphibole is absent.

In the Antilles Island Arc, phenocrysts up to An_{85} occur in andesite and basalt from Montserrat, along with amphibole-plagioclase (An_{93-89})-magnetite xenoliths⁷⁰. At Soufrière Volcano, St Vincent, ejected plutonic blocks contain plagioclase (An_{96-89}), olivine (For_{79-67}), salite, hastingsitic amphibole and Mg–Al–Ti-magnetite⁷¹. The Soufrière amphibole has too little SiO_2 , the magnetite has no detectable Cr, and the salite has too much Al_2O_3 to match the Fiskenaesset minerals. The Salt Lick Creek layered intrusion of Western Australia⁷² contains plagioclase (An_{84}) associated with olivine (For_{81-84}), orthopyroxene (Fs_{15-20}), clinopyroxene ($Wo_{48-44}En_{46-48}Fs_{6-8}$) and Mg–Al-chromite, in assemblages similar to those in layered complexes in Rhum⁷³, the Cuillins of Skye⁷⁴, and Mull⁷⁵. The absence of amphibole and magnetite distinguishes this suite from the Fiskenaesset Complex. Megacrysts of An_{87-93} with inclusions of augite or perhaps augite plus amphibole occur in feldspathic dykes of allivalite composition in Skye⁷⁶, and ultrabasic breccias in the Rhum Complex⁶⁴ contain plagioclase (An_{87-97}), olivine (For_{83-82}), augite ($Wo_{43-46}En_{46-48}Fs_{8-9}$; 4 weight % Al_2O_3), primary kaersutite (TiO_2 , 4.9; Al_2O_3 , 11.6; Cr_2O_3 , 0.5; FeO, 8.2; CaO, 11.9; Na_2O , 2.6; K_2O , 1.0 weight %), phlogopite, Mg–Al–Ti chromite–magnetite and picroilmenite. Although the amphibole is too high in TiO_2 and low in SiO_2 and the augite is too rich in Al_2O_3 , this mineral assemblage might resemble selected assemblages in the Fiskenaesset Complex after metamorphism of the former at about 5 kbar and 900 °C. The Greenhills Ultramafic Complex of New Zealand⁷⁷, ranges from chromite-bearing dunite to wehrlite, gabbro and allivalite, all three of which carry titanomagnetite and brown hornblende. The sequence of the opaque minerals provides a ready distinction from the Fiskenaesset Complex, but major-element compositions of the olivine, pyroxene and plagioclase (An_{88-92}) lie in the Fiskenaesset ranges.

Alpine-type ultramafic complexes⁷⁸ can be dismissed because their olivines and orthopyroxenes are too magnesian rich (mg near 0.9) to match the range (0.86–0.74) for the Fiskenaesset Complex. Concentrically-zoned ultramafic complexes of Alaskan type^{52,78,79} contain rocks with various amounts of primary hornblende, olivine, clinopyroxene, magnetite, chromite, ilmenite and hercynite spinel with compositions fairly similar to those of the Fiskenaesset minerals; but plagioclase is virtually absent except in late hornblende-plagioclase pegmatite, and the type of zoning differs from that at Fiskenaesset. The interpretation³⁶ of these zoned complexes as feeder pipes of andesitic volcanoes is feasible, and the resemblance of their mineral compositions to those at Fiskenaesset could be attributable to the crystallisation of a hydrous basaltic magma in

both cases, one being open and the other closed. Layered gabbroic and ultramafic rocks in a tectonic melange on the Kodiak islands of Alaska have received only a brief description⁸⁰, but they should be compared closely with Archaean complexes. Ultramafic rocks at Emigrant Gap, California⁸¹, are part of a mafic complex showing near-concentric zoning from dominant wehrlitic peridotite in the core to tonalite in the rim. Olivine, pyroxenes and plagioclase from the inner rocks lie in the composition range of the Fiskenaesset rocks. Magnetite is ubiquitous and chromite also occurs, but no compositions have been determined. Amphibole is absent in many rocks though it occasionally replaces augite. Detailed chemical studies are needed, but the differentiation trend to tonalite apparently results because the water content was too low to produce substantial amounts of amphibole.

The Fiskenaesset Complex fits best with the numerous bodies (~ 100) of peridotite-gabbro of up to 70 square kilometres in outcrop dispersed among the larger tonalite-granodiorite bodies of the Peninsular Range Batholith of southern California^{82,83}. The mineral assemblages and compositions of the peridotites, gabbros and anorthositic gabbros⁸⁵ can be matched with those at Fiskenaesset except for small details, and the interpretation in terms of the differentiation of a wet basaltic–andesitic magma is similar. The differences between the two mineralogies are explicable in terms of different degrees of metamorphism and small variations in the differentiation trend resulting from small changes of water pressure, oxidation state, and so on.

Although more detailed mineralogical petrological–geochemical studies are needed of layered complexes in Archaean and modern environments, our survey indicates that, first, Archaean complexes can be explained by the metamorphism of igneous complexes developed from wet basaltic magmas differentiating under oxidising conditions, and, second, the Archaean complexes can be matched chemically with the complexes in the southern California batholith. Mineralogical data for the layered complexes of the Caledonide fold belt of Ireland and Scotland are sparse (refs 37–39, 84 and 85) but the presence of calcic plagioclase, apparently primary hornblende, and medium-Cr, medium-Fe spinel are tantalising clues.

Discussion

The crux of our argument is that the late Archaean intrusive tonalites (now tonalitic gneisses and granulites) differentiated from a more basic, mantle-derived magma generated in some type of active continental margin. We suggest that enclosed, layered igneous complexes can be matched mineralogically with complexes enclosed within a tonalitic batholith of western North America, and that the associated Archaean and Mesozoic orthoquartzite-carbonate-K-pelite associations are both typical of a shallow-water continental margin. Furthermore, our argument that the Archaean high grade regions are equivalent to the deep-seated levels of modern arcs is consistent with the current interpretation of greenstone belts as fossil back-arc marginal basins^{86,87}.

Because Archaean tectonic conditions and magma compositions may differ from present ones, analogies must not be pressed too far. Whereas the granitic diapirs and vertical, near-cylindrical, zoned complexes of the Alaskan type reflect vertical transport in modern continental margins, strong arguments favour horizontal tectonics in the Archaean¹⁰. The apparently greater abundance of komatiitic lavas in Archaean times can be explained by a steeper geothermal gradient resulting from higher heat production by younger radioactive species. Early basaltic magmas may have been wetter than younger ones because the Earth was undergoing strong outgassing in the Archaean (see, for example, ref. 88). Early plates may have been smaller, less coherent and more mobile. Nevertheless, we suggest that the growth of continents in the late Archaean can be studied profitably in the context of models related to accretion at the leading edge of modern continental plates.

B. F. W. thanks the Geological Survey of Greenland and the NERC and J. V. S. thanks F. C. Bishop, I. M. Steele, NASA and the NSF. R. K. Nishimori and J. Hawkins contributed discussions and we thank P. J. Wyllie and R. C. Newton for reviews.

- ¹ Bridgwater, D., Watson, J. V., and Windley, B. F., *Phil. Trans. R. Soc.*, **A273**, 493 (1973).
- ² Kalsbeek, F., in *Early History of the Earth* (edit. by Windley, B. F.) (Wiley, London, in the press).
- ³ Bowes, D. R., Barooah, B. C., and Khoury, S. G., *Spec. Publ. geol. Soc. Aust.*, **3**, 77 (1971).
- ⁴ Viswanathan, S., *J. Geol. Soc. Ind.*, **15**, 347 (1974).
- ⁵ Barker, F., and Peterman, Z. E., *Precamb. Res.*, **1**, 1 (1974).
- ⁶ Sheraton, J. W., *Earthplanet. Sci. Lett.*, **8**, 301 (1970).
- ⁷ Tarney, J., Skinner, A. C., and Sheraton, J. W., *Proc. 24th int. Geol. Congr., Montreal*, **1**, 162 (1972).
- ⁸ Holland, J. G., and Lambert, R. St. J., *Precamb. Res.*, **2**, 161 (1975).
- ⁹ McGregor, V. R., *Phil. Trans. R. Soc.*, **A273**, 343 (1973).
- ¹⁰ Bridgwater, D., McGregor, V. R., and Myers, J. S., *Precamb. Res.*, **1**, 179 (1974).
- ¹¹ Tarney, J., in *Early History of the Earth* (edit. by Windley, B. F.) (Wiley, London, in the press).
- ¹² Windley, B. F., *Spec. Publ. geol. Soc. S. Afr.*, **3**, 312 (1973).
- ¹³ Watson, J. V., in *Spec. Rep. geol. Soc. Lond.*, **6**, 15 (1975).
- ¹⁴ Collerson, K. D., Jesseau, C. W., and Bridgwater, D., in *Early History of the Earth* (edit. by Windley, B. F.) (Wiley, London, in the press).
- ¹⁵ Mason, R., *Phil. Trans. R. Soc.*, **A273**, 463 (1973).
- ¹⁶ Radhakrishna, B. P., *J. geol. Soc. Ind.*, **15**, 439 (1974).
- ¹⁷ Moorbatch, S., *Nature*, **254**, 395 (1975).
- ¹⁸ Hor, A. K., Hutt, D. K., Smith, J. V., Wakefield, J., and Windley, B. F., *Lithos*, **8**, 297 (1975).
- ¹⁹ Myers, J. S., in *Rapp. Gronlands Geol. Unders.*, **51**, 47 (1973).
- ²⁰ Windley, B. F., Herd, R. K., and Bowden, A. A., *Bull. Gronlands Geol. Unders.*, **106** (1973).
- ²¹ Myers, J. S., in *Early History of the Earth* (edit. by Windley, B. F.) (Wiley, London, in the press).
- ²² Windley, B. F., and Selvan, T. A., *J. geol. Soc. Ind.*, **16**, 209 (1975).
- ²³ Windley, B. F., *Mem. Ass. Am. Petrol. Soc.*, **12**, 899 (1969).
- ²⁴ Garson, M. S., and Livingstone, A., *Nature phys. Sci.*, **243**, 74 (1973).
- ²⁵ Engel, C. G., and Fisher, R. L., *Bull. geol. Soc. Am.*, **86**, 1553 (1975).
- ²⁶ Sylvester-Bradley, P. C., *Proc. R. Soc.*, **B189**, 213 (1975).
- ²⁷ Sutton, J., in *Early History of the Earth* (edit. by Windley, B. F.) (Wiley, London, in the press).
- ²⁸ Gunn, B. M., in *Early History of the Earth* (edit. by Windley, B. F.) (Wiley, London, in the press).
- ²⁹ Rivalenti, G., in *Early History of the Earth* (edit. by Windley, B. F.) (Wiley, London, in the press).
- ³⁰ Miyashiro, A., *Am. J. Sci.*, **274**, 321 (1974).
- ³¹ Brooks, C., and Hart, S. R., *Geology*, **2**, 107 (1975).
- ³² Carmichael, I. S. E., Turner, F. J., and Verhoogen, J., *Igneous Petrology* (McGraw-Hill, New York, 1974).
- ³³ Reed, B. L., and Lanphere, M. A., *J. Res. U.S. geol. Surv.*, **2**, 343 (1974).
- ³⁴ Brown, G. C., *Nature phys. Sci.*, **241**, 26 (1973).
- ³⁵ Nishimori, R. K., *Geol. Soc. Am., Abstr. with Progr.*, **6**, 229 (1974).
- ³⁶ Murray, C. G., *Mem. geol. Soc. Am.*, **132**, 313 (1972).
- ³⁷ Leake, B. E., in *Mechanism of Igneous Intrusion* (edit. by Newall, G., and Rast, N.), *Geol. J. Liverpool, Spec. Issue*, **2**, 103 (1970).
- ³⁸ Rothstein, A. T. V., *Q. J. geol. Soc. Lond.*, **113**, 1 (1957).
- ³⁹ Leake, B. E., *Geol. Mag.*, **101**, 63 (1964).
- ⁴⁰ Leake, B. E., *Q. J. geol. Soc. Lond.*, **125**, 219 (1970).
- ⁴¹ Dewey, J. F., in *The Ocean Basins and Margins*, **2** (edit. by Nairn, A. E. M., and Stehli, F. G.), **205** (Plenum, New York, 1975).
- ⁴² *The Geology of Scotland* (Oliver and Boyd, Edinburgh, 1965).
- ⁴³ Windley, B. F., and Smith, J. V., *Bull. Gronlands Geol. Unders.*, **108** (1974).
- ⁴⁴ Subramaniam, A. P., *Bull. geol. Soc. Am.*, **67**, 317 (1956).
- ⁴⁵ Dearnley, R., *Q. J. geol. Soc. Lond.*, **119**, 243 (1963).
- ⁴⁶ Wood, B. J., *Earthplanet. Sci. Lett.*, **26**, 299 (1975).
- ⁴⁷ Chappell, B. W., and White, A. J. R., *Mineralog. Mag.*, **37**, 555 (1970).
- ⁴⁸ Eggler, D. H., *Earthplanet. Sci. Lett.*, **15**, 28 (1972).
- ⁴⁹ Helz, R. T., *J. Petrol.*, **14**, 249 (1973).
- ⁵⁰ Cawthorn, R. G., Curran, E. B., and Arculus, R. J., *J. Petrol.*, **14**, 327 (1973).
- ⁵¹ Mysen, B. O., and Boettcher, A. L., *J. Petrol.*, **16**, 549 (1975).
- ⁵² Irvine, T. N., *Mem. geol. Soc. Am.*, **138** (1974).
- ⁵³ Rodgers, K. A., and Bevan, J. C., *Mineralog. Mag.*, **39**, 890 (1974).
- ⁵⁴ Donaldson, C. H., *J. Geol.*, **83**, 33 (1975).
- ⁵⁵ Osborn, E. F., *Am. Miner.*, **47**, 211 (1962).
- ⁵⁶ Hill, R., and Roeder, P., *J. Geol.*, **82**, 709 (1974).
- ⁵⁷ Henderson, P., *Geochim. cosmochim. Acta*, **39**, 1035 (1975).
- ⁵⁸ Cawthorn, R. G., and O'Hara, M. J., *Am. J. Sci.* (in the press).
- ⁵⁹ Cawthorn, R. G., *J. Petrol.* (in the press).
- ⁶⁰ Green, T. H., and Ringwood, A. E., *Contr. Miner. Petrol.*, **18**, 105 (1968).
- ⁶¹ Yoder, H. S., Jr., and Tilley, C. E., *J. Petrol.*, **3**, 342 (1962).
- ⁶² Rodgers, K. A., and Brothers, R. N., *Mineralog. Mag.*, **37**, 375 (1969).
- ⁶³ White, R. W., *Contr. Miner. Petrol.*, **12**, 254 (1966).
- ⁶⁴ Aoki, K., and Shiba, I., *Sci. Rep. Tohoku Univ., Ser. III*, **12**, 395 (1974).
- ⁶⁵ Scheidegger, K., *J. geophys. Res.*, **78**, 3340 (1973).
- ⁶⁶ Bougault, H., and Hekinian, R., *Earthplanet. Sci. Lett.*, **24**, 249 (1974).
- ⁶⁷ Baxter, A. N., *Bull. geol. Soc. Am.*, **86**, 1449 (1975).
- ⁶⁸ Ewart, A., Bryan, W. B., and Gill, J. B., *J. Petrol.*, **14**, 429 (1973).
- ⁶⁹ Blake, D. H., and Ewart, A., *J. geol. Soc. Aust.*, **21**, 319 (1974).
- ⁷⁰ Rea, W. J., *J. geol. Soc. Lond.*, **130**, 341 (1975).
- ⁷¹ Lewis, J. F., *Contr. Miner. Petrol.*, **38**, 197 (1973).
- ⁷² Wilkinson, J. F. G., Duggan, M. B., Herbert, H. K., and Kalocsai, G. I. Z., *Contr. Miner. Petrol.*, **50**, 1 (1975).
- ⁷³ Brown, G. M., *Phil. Trans. R. Soc.*, **B240**, 1 (1956).
- ⁷⁴ Weedon, D. S., *Scott. J. Geol.*, **1**, 41 (1965).
- ⁷⁵ Wager, L. R., and Brown, G. M., *Layered Igneous Rocks* (Oliver and Boyd, London, 1968).
- ⁷⁶ Donaldson, C. H., *Earthplanet. Sci. Lett.*, **27**, 251 (1975).
- ⁷⁷ Mossman, D. J., *Bull. geol. Soc. Am.*, **84**, 39 (1973).
- ⁷⁸ Wyllie, P. J. (ed.), *Ultramafic and Related Rocks* (Wiley, New York, 1967).
- ⁷⁹ Ruckmick, J. C., and Noble, J. A., *Bull. geol. Soc. Am.*, **70**, 981 (1959).
- ⁸⁰ Hill, B. B., *Geol. Soc. Am., Abstr. with Progr.*, **7**, 1116 (1975).
- ⁸¹ James, O. B., *J. Petrol.*, **12**, 523 (1971).
- ⁸² Miller, F. S., *Bull. geol. Soc. Am.*, **48**, 1397 (1937).
- ⁸³ Gastil, R. G., *Geology*, **3**, 361 (1975).
- ⁸⁴ Rothstein, A. T. V., *Geol. Mag.*, **95**, 456 (1958).
- ⁸⁵ Rothstein, A. T. V., *Mineralog. Mag.*, **38**, 957 (1972).
- ⁸⁶ Tarney, J., Dalziel, I. W. D., and de Wit, M. J., in *Early History of the Earth* (edit. by Windley, B. F.) (Wiley, London, in the press).
- ⁸⁷ Burke, K., Dewey, J. F., and Kidd, W. S. F., in *Early History of the Earth* (edit. by Windley, B. F.) (Wiley, London, in the press).
- ⁸⁸ Rubey, W. W., *Spec. Pap. geol. Soc. Am.*, **62**, 631 (1955).

articles

Observation of light emission from a rhodopsin

Aaron Lewis, John P. Spoonhower & Gregory J. Perreault

School of Applied and Engineering Physics, Cornell University, Ithaca, New York 14853

Radiative emission has been observed from the retinylidene chromophore of bacteriorhodopsin, the purple membrane protein of Halobacterium halobium.

LIGHT emission has never been observed from the retinylidene chromophore of any rhodopsin¹. In addition, no vibronic structure has been detected in the absorption spectrum of the chromophore of rhodopsin or of any isomer of retinal. Furthermore, no reproducible electron paramagnetic resonance signal has been recorded from any rhodopsin. These are the principal reasons why little is known about the electronic structure of the retinylidene chromophore of rhodopsin.

Theoretical calculations² and studies of model polyenes^{3,4}

have been the principal approaches used to try to understand the electronic structure of various rhodopsins. We report here the first observation of light emission attributable to the retinylidene chromophore of a rhodopsin. This study and several other investigations in progress in our laboratory are directed toward resolving the excited state ordering in the retinylidene chromophore and understanding changes in these excited states after a photon interacts with rhodopsin.

Emission spectroscopy of rhodopsin

While applying resonance Raman spectroscopy to study the ground state vibrational structure of the retinylidene chromophore of several rhodopsins and their intermediates⁵⁻¹⁰, we

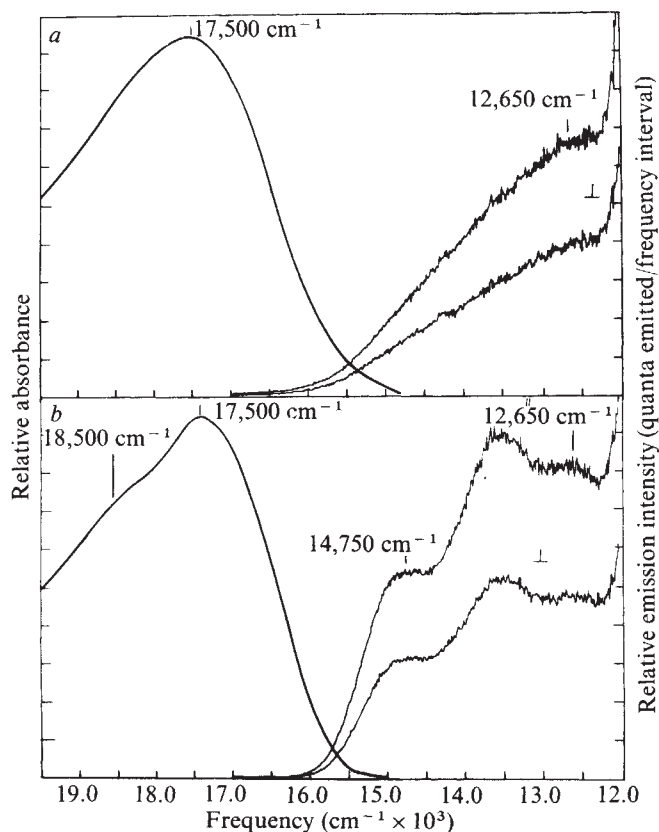


Fig. 1 *a*, Room temperature relative absorbance in arbitrary units scaled from ref. 21 and room temperature relative emission intensity, corrected for instrument response in units of number of quanta emitted per unit frequency interval. The ordinate scale is in thousands of wave numbers. The excitation source for the emission is the 514.5-nm line of an argon ion laser and intensities both parallel and perpendicular to the incident laser polarisation are shown. The apparent rapid rise in the extreme low energy portion of the corrected spectra is artificial and due to division of a nearly constant dark count for the photomultiplier by the vanishing instrument response function. *b*, Low temperature (77 K) relative absorbance in arbitrary units scaled from ref. 21 and 77 K emission intensity corrected for instrument response in units of number of quanta emitted per unit frequency interval. The ordinate scale is in thousands of wavenumbers. The excitation source for the emission is the 580.0-nm line of a CW rhodamine 6G dye laser. Intensities both parallel and perpendicular to the incident laser polarisation are shown.

have developed, as a by-product, a spectrometer which is also a sensitive fluorimeter. The excitation source for our spectrometer is a tunable dye laser, which provides an ideal source for the stimulation of weak emissions.

We have used our Raman spectrometer to detect emission from rhodopsins and have observed re-emission from bacteriorhodopsin. Bacteriorhodopsin has been studied by Stoekenius *et al.*^{11–14}, who have shown that *Halobacterium halobium* uses this rhodopsin to convert light energy into chemical energy. Bacteriorhodopsin has an absorption maximum at 570 nm, and when the retinylidene chromophore is extracted from this species it has an all-*trans* conformation¹⁵. Bacteriorhodopsin occurs in purple membrane patches in the plasma membrane of the bacteria. These patches can be isolated as membrane fragments which contain the rhodopsin in a crystalline hexagonal array, as shown by diffraction experiments^{12,16}. The array of rhodopsins is in clusters of three in the membrane with a threefold axis of symmetry perpendicular to the plane of the membrane at the centre of each cluster¹⁶. We have used bacteriorhodopsin prepared by the procedure of Oesterhelt and Stoekenius¹⁷. The membrane fragments were suspended in water at room temperature and in a glycerol-water glass at 77 K. Similar results were obtained from

membrane fragments isolated from an S-9 mutant of the bacteria, which is believed not to contain carotenoids.

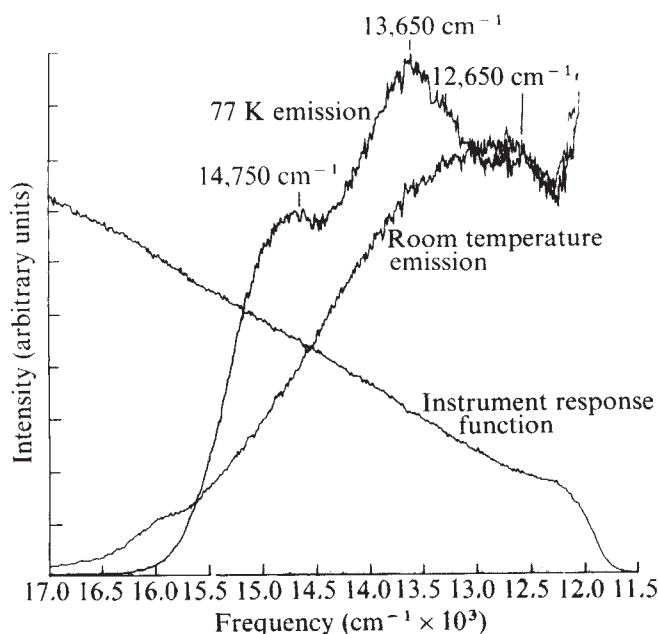
All the spectral data shown in Figs 1–3 were obtained on a laser Raman spectrometer using a red sensitive RCA C31034 photomultiplier with photon counting electronics. Each data point was scaled digitally with the measured instrument response function as determined using a quartz-halogen standard lamp. The instrument response function can be seen in Fig. 2 superimposed on the room temperature and 77 K corrected spectra. The excitation wavelengths used for these measurements were obtained from either an argon ion laser or a continuous wave rhodamine 6G tunable dye laser. All absorption spectra were recorded on a Cary 14 spectrophotometer.

Quantum efficiency

Bacteriorhodopsin is ideal for the study of re-emission from rhodopsins because the membrane fragments contain only one protein^{12,16}, which is the rhodopsin. The emission maximum we observed at room temperature (Fig. 1*a*) occurs in the near infrared region of the spectrum and is shifted by ~4,850 cm⁻¹ from the room temperature absorption maximum. There is minimal overlap between the absorption and emission spectra at room temperature and at 77 K. The room temperature quantum efficiency was determined to be between 1.2 (10⁻⁴) and 2.5 (10⁻⁴), using reduced photosynthetic reaction centres from the bacterium *Rhodospseudomonas spheroides* as a standard, according to the method of Parker and Rees¹⁸. A quantum efficiency of 1.0 (10⁻³) was assumed for the reaction centres¹⁹.

Reaction centres are an ideal standard for this study because of their relatively low emission quantum yield and because of the high degree of spectral overlap with the emission of bacteriorhodopsin. Absorbance spectra, taken before and after the measurement of reaction centre emission, indicated no changes in the preparation. The absorbance of bacteriorhodopsin used for the quantum yield was 0.687, while the reaction centres had an absorbance of 0.585. The excitation source for just the quantum efficiency measurements was a quartz-halogen lamp in combination with a Corning 4-96 filter and a broad band 550-nm interference filter. The instrument response function was determined with a lamp having a

Fig. 2 Relative instrument response superimposed on the room temperature and 77 K corrected emission spectra. The emission spectra are in units of number of quanta emitted per unit frequency interval. The band observed at approximately 16,000 cm⁻¹ in the room temperature emission spectrum is due to the ~3,400-cm⁻¹ band in the Raman spectrum of H₂O.



colour temperature of 2,860 K. The spectra were recorded in an analogue mode with a photomultiplier having an S-1 response and were subsequently digitised to allow for a point-by-point subtraction of background due to scatter. Scattering spectra were obtained using a Ludox solution. Furthermore, the spectra were scaled by the instrument response function and corrected to give the integrated area in terms of photons emitted per incident photon. No effect on emission intensity was observed due to deaeration. A detailed study of the quantum efficiency in various conditions will be reported elsewhere.

Polarisation of the emission

The emission intensities of the components parallel and perpendicular to the incident laser polarisation were determined using a Pockels cell to rotate the laser polarisation with respect to a fixed analyser. We recorded the parallel and perpendicular emission components in alternate channels at each frequency setting of the monochromator. The data were separated, scaled and plotted using a digital computer and are presented in Fig. 1a and b. The emission polarisation $[(I_{||} - I_{\perp}) / (I_{||} + I_{\perp})]$ which we have measured in a 90° scattering geometry varies from 0.23 ± 0.02 to 0.49 ± 0.02 at both room temperature and 77 K when excitation frequencies of $19,436 \text{ cm}^{-1}$ and $17,241 \text{ cm}^{-1}$ are used. Such variations in the degree of polarisation have been observed for pigments in *Chlorella*, for example²⁰. These variations when different experiments are compared result from varying amounts of partial orientation in the gravitational field of the suspension of membrane fragments containing the well-ordered pigment molecules. Because of this effect, one way to determine the degree of emission polarisation accurately is to measure it in the presence of a magnetic field: this is in progress in our laboratory. It is interesting, however, that the degree of polarisation seen in Fig. 1a and b remains constant throughout the observed emission band within the time scale of any one measurement. Thus these data suggest that the emission emanates from the excited state of a single species.

This conclusion is supported by a plot of the ratio of the intensities of the components at $12,650 \text{ cm}^{-1}$, $13,650 \text{ cm}^{-1}$ and $14,750 \text{ cm}^{-1}$ in the 77 K spectrum as a function of exciting laser frequency. The data plotted in Fig. 3 are for the exciting laser frequencies which produce the largest changes in the concentration of the photoproduct (unpublished results of R. Lozier and W. Stoekenius). (The photoproduct at 77 K is a red-shifted intermediate with an absorption maximum $\sim 635 \text{ nm}$; refs. 21 and 22). Therefore these experiments also indicate that the components observed in the 77 K spectrum arise from a single species, and this suggests that the components are indeed vibronic fine structure on the emission. The spacing between the components is $\sim 1,000 \text{ cm}^{-1}$ and $\sim 1,100 \text{ cm}^{-1}$ in the emission spectrum, and a shoulder is detected at a similar spacing on the high energy side of the 77 K absorption maximum (Fig. 1b).

Assignment of observed emission

Light induces photochemical changes in bacteriorhodopsin similar to those observed in all rhodopsins. This photochemical behaviour changes the concentration of rhodopsin with different actinic wavelengths. A two-laser-beam technique can be used to control the photochemical behaviour while observing the emission spectrum. This is particularly feasible at 77 K, where only two components are present, bacteriorhodopsin and the batho form absorbing at $\sim 635 \text{ nm}$ (refs 22 and 23). These two components are photoreversible at 77 K and exist in a photostationary equilibrium depending on actinic wavelength (unpublished results of R. Lozier and W. Stoekenius). We have obtained data by the use of two actinic wavelengths. The data are summarised in Table 1. The principal feature of these experiments is similar to the enhancement effects observed in photosynthetic systems, where the effect of two simultaneous actinic irradiations is not equal to the sum of the individual

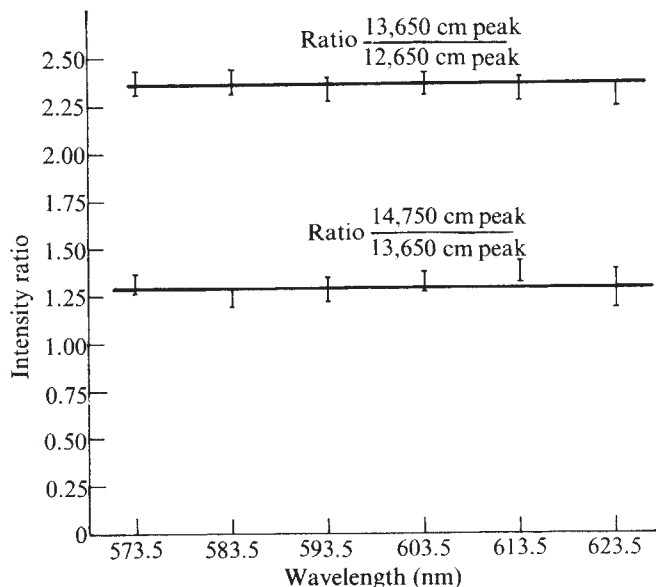


Fig. 3 Ratios of intensities of the bands observed at $14,750 \text{ cm}^{-1}$, $12,650 \text{ cm}^{-1}$ and $13,650 \text{ cm}^{-1}$ against wavelength. The excitation source is a continuous wave rhodamine 6G dye laser. The ratios shown are calculated from unscaled data.

effects^{24,25}. The sum of the emission intensities stimulated by 10 mW of illumination at 636.0 nm and by 1 mW of illumination at 514.5 nm obtained in separate experiments is less than the emission observed when the sample is irradiated simultaneously with both beams. In addition, the sum of emissions stimulated with 1 mW of 636.0 nm and 10 mW of 514.5 nm is less than the emission obtained in a single dual beam experiment with the same laser powers and wavelengths. The differences we observed can be explained in terms of changes in the photostationary concentrations of bacteriorhodopsin and the photochemically produced batho form. This is independent of any assumption about the wavelength dependence of the quantum efficiency of the emission and the photochemical quantum yield. The reason for this is that, in the experiments tabulated in Table 1, only the laser powers and not the wavelengths have been changed.

Table 1 is consistent with the interpretation that the primary emitter, in the region we detected the emission spectrum, is bacteriorhodopsin. For example, the addition of 1 mW at 636.0 nm in experiment 6 not only adds its own emission intensity but also increases the steady state concentration of bacteriorhodopsin to a value which is greater than that of experiment 5. Thus the intensity observed in experiment 6 is greater than the sum of the intensities observed in experiments 4 and 5. A similar explanation can be made for experiments 1, 2 and 3.

A comparison of the room temperature and 77 K emission

Table 1 Emission intensities observed using a dual beam technique

Experiment	Laser I (636.0 nm) (mW)	Laser II (514.5 nm) (mW)	Emission intensity proportional to (quanta emitted/ incident quanta)
1	10	0	267
2	0	1	69
3	10	1	386
4	1	0	26
5	0	10	694
6	1	10	738

All dual beam experiments make use of both an argon ion laser to produce radiation at 514.5 nm and a CW rhodamine 6G dye laser for 636.0-nm irradiation. The emission intensities have been corrected so that they are given in relative number of quanta emitted per number of incident quanta. Each emission intensity is accurate to $\pm 2\%$.

spectra (Fig. 3) is also consistent with the suggestion that bacteriorhodopsin is the primary emitter. It is known that, based on the kinetic constant at room temperature for the decay of the batho intermediate to its thermal product^{22,23}, appreciable concentrations of this intermediate cannot be present at room temperature. Conversely, at the temperature of liquid nitrogen a significant fraction of the photostationary mixture is in the batho form. When this is considered together with the similarities in the room temperature and 77 K spectra, with the observation that the ratio of the vibronic components remains constant at various exciting wavelengths (Fig. 3) and with the results of the two beam experiments, we conclude that the emission emanates from bacteriorhodopsin and that the batho form does not emit in this region of the spectrum. This conclusion is supported by excitation spectra which we have obtained at room temperature.

The excited state

Our experiments indicate that an excited state of bacteriorhodopsin is responsible for the emission spectra we observed. Based on the integrate strength of the absorption for the allowed transition, a radiative lifetime for emission from the 1B_u state would be 7 ns. But measurements of the lifetime by the use of a mode-locked tunable dye laser and by phase-modulation techniques suggest that the calculated intrinsic lifetime of the emission is between 0.1 and 1 μ s (our unpublished results with M. Marcus, M. Hirsch and A. Kriebel). Thus the emission cannot originate from the 1B_u state. The observed variation in the lifetime of the emission could result from exciton interactions among the chromophores at high laser intensities.

There is, however, a separation between the absorption and emission bands (Fig. 1a and b) similar to the separation observed by Hudson and Kohler⁴ in the spectra of model polyenes; the radiative lifetimes we observe are comparable with those detected by Hudson and Kohler²⁶ for model polyenes. The data on these model systems have been taken to mean that the lowest energy excited state in these polyenes is a 1A_g state. The occurrence of such a perfect 1A_g state, or a perfect 1B_u allowed state, in the retinylidene chromophore is unlikely especially in view of the large excited state dipoles that are measured (personal communication from R. Mathies and L. Stryer) and the observation of a circular dichroism spectrum in bacteriorhodopsin²⁷. Thus our data can only be interpreted to mean that the emission emanates from some state other than the allowed state into which light energy is absorbed. Picosecond absorption measurements have suggested the existence of a transient intermediate produced before the batho intermediate which occurs in < 6 ps (ref. 28). This transient could represent the absorption of light by the state from which we are observing emission.

We can deduce from our data some information on the character of the excited state from which emission is observed. For example, the observation of vibronic structure in the 77 K emission spectrum suggests that the excited state involved in the emission has electron density delocalised in the polyene. This would increase the bond order of single bonds and decrease the bond order of double bonds. It has been suggested² that a lack of vibronic structure in the absorption spectrum can be attributed to the sterically hindered cyclohexene ring in the retinylidene chromophore. Thus the C6-C7 single bond between the cyclohexene ring and the isoprenoid chain has a very broad ground-state potential. But if in the excited state the electron density is delocalised, the bond order of the C6-C7 single bond should increase and this in turn should narrow considerably the excited state potential surface for this torsional angle coordinate. Such an effect would reduce the conformational freedom of the cyclohexene ring relative to the isoprenoid chain and this should result in the vibronic structure observed in our emission spectrum. We believe it is significant that resonance Raman experiments on the batho intermediate (our unpublished results) which is stabilised from the excited

state also indicate (from the frequencies of the C-CH₃, -C=N- and C=C vibrations) significant reduction in the C=C bond order and charge polarisation in this first stable product which results from a photon interacting with bacteriorhodopsin. Thus when the fluorescence and resonance Raman data are analysed together the excited state structure and electronic distribution seem to be a precursor and initiator of the stable batho photochemical product.

The blueshift we observed in the emission spectra when changing from room temperature to 77 K suggests that the chromophore is surrounded in the lipoprotein matrix by dipolar groups which have not relaxed fully at low temperature. In addition it is clear from the two beam experiments completed at 77 K that the batho form does not have the same excited state as this rhodopsin. Some theories of the photochemistry of rhodopsin have predicted that the excited states of rhodopsin and the batho intermediate are identical. Finally, the fact that no emission is observed for the batho form in the region of the spectrum where our apparatus is sensitive strongly suggests that the photochemical act which stores light energy in the 635-nm intermediate quenches radiative emission.

In conclusion, we feel that these observations of emission from a rhodopsin open new opportunities to understand the excited state ordering and structure of the retinylidene chromophore. Furthermore, these emission spectra will be important in understanding the primary photonic event which produces a red shift in the chromophores of all rhodopsins. Thus these data on the initial photochemically induced changes in the retinylidene chromophore should yield new insights into the conversion of light energy into chemical energy by rhodopsin, and should have an important role in the formulation of a mechanism of energy transduction by rhodopsin.

We thank Professor Walther Stoeckenius for bacteriorhodopsin and helpful discussions. We also thank Dr Richard Lozier for stimulating discussions; Professors A. Albrecht and H. Mahr, Dr A. Kriebel, and M. Hirsch for obtaining the emission lifetime, and Professor R. Clayton for encouragement and obtaining the room temperature quantum efficiency. A.L. is an Alfred P. Sloan fellow; J.P.S. held an Eastman Kodak graduate fellowship; G.J.P. was supported by a graduate fellowship from the Ford Foundation. This work was supported by a grant from the National Institutes of Health.

Received August 11, 1975; accepted February 27, 1976.

- 1 Busch, G. E., Applebury, M. L., Lamola, A. A., and Rentzepis, P. M., *Proc. natn. Acad. Sci. U.S.A.*, **69**, 2802-2806 (1972).
- 2 Honig, B., Warshel, A., and Karplus, M., *Acc. chem. Res.*, **8**, 92-100 (1975).
- 3 Moore, T. A., and Song, P., *Chem. Phys. Lett.*, **19**, 128-130 (1973); Gavin, R. M., Risenberg, S., and Rice, S. A., *J. Chem. Phys.*, **58**, 3160-3165 (1973).
- 4 Hudson, B. S., and Kohler, B. E., *A. Rev. Phys. Chem.*, **25**, 437-460 (1974).
- 5 Lewis, A., Fager, R., and Abrahamson, E. W., *J. Raman Spectrosc.*, **1**, 465-470 (1973).
- 6 Lewis, A., and Spoonhower, J., in *Spectroscopy in Biology and Chemistry*, Ch. II, 347-376 (edit. by Chen, S., and Yip, S.) (Academic, New York, 1974).
- 7 Lewis, A., Spoonhower, J., Bogomolnoi, R., Lozier, R., and Stoeckenius, W., *Proc. natn. Acad. Sci. U.S.A.*, **71**, 4462-4466 (1974).
- 8 Lewis, A., *Biophys. J.*, **15**, 174 (1975).
- 9 Lewis, A., *Fedn Proc.*, **35**, 51-53 (1976).
- 10 Cockingham, R. E., Lewis, A., Collins, D. W., and Marcus, M., *J. Am. Chem. Soc.* (in the press).
- 11 Stoeckenius, W., and Oesterheld, D., *Nature new Biol.*, **233**, 149-152 (1971).
- 12 Blaurock, A., and Stoeckenius, W., *Nature new Biol.*, **233**, 152-154 (1971).
- 13 Oesterheld, D., and Stoeckenius, W., *Proc. natn. Acad. Sci. U.S.A.*, **70**, 2853-2857 (1973).
- 14 Racker, E., and Stoeckenius, W., *J. biol. Chem.*, **249**, 662-663 (1974).
- 15 Oesterheld, D., *Eur. J. Biochem.*, **40**, 453-463 (1973).
- 16 Henderson, R., and Unwin, P. N. T., *Nature*, **257**, 28-32 (1975).
- 17 Oesterheld, D., and Stoeckenius, W., *Meth. Enzymol. Biomembranes*, **31**, Part A, 667-678 (1974).
- 18 Parker, C. A., and Rees, W. T., *Analyst*, **85**, 587-600 (1960).
- 19 Zanker, K. L., Reed, D. W., and Clayton, R. K., *Proc. natn. Acad. Sci. U.S.A.*, **61**, 1243-1249 (1968).
- 20 Breton, J., Becker, J. F., and Geacintov, N. E., *Biochem. biophys. Res. Commun.*, **54**, 1403 (1973).
- 21 Stoeckenius, W., and Lozier, R., *J. supramolec. Struct.*, **2**, 769-774 (1974).
- 22 Chu King, M., DeVault, D., Hess, B., and Oesterheld, D., *Biophys. J.*, **15**, 907-911 (1975).
- 23 Lozier, R., and Stoeckenius, W., *Biophys. J.*, **15**, 955 (1975).
- 24 Emerson, R., Chalmers, R., and Cederstrand, C., *Proc. natn. Acad. Sci. U.S.A.*, **43**, 133-143 (1957).
- 25 Butler, W. L., *Acc. chem. Res.*, **6**, 177-184 (1973).
- 26 Hudson, B. S., and Kohler, B. E., *J. chem. Phys.*, **59**, 4984-5002 (1973).
- 27 Beecher, B., and Cassim, J. Y., *Biophys. Soc. Abstr.*, **65** (1975).
- 28 Kaufmann, K., Rentzepis, P. M., and Lewis, A., *Proc. natn. Acad. Sci. U.S.A.* (in the press).

Bicycle-pedal model for the first step in the vision process

Arieh Warshel*

Department of Chemical Physics, Weizmann Institute of Science, Rehovot, Israel, and MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Computer simulation of the molecular dynamics of retinal during its photoisomerisation inside a restrictive active site gives a detailed model for the sequence of events in the first step of the vision process. It is proposed that the prelumirhodopsin intermediate contains a strained all-trans retinal molecule produced directly and rapidly from the 11-cis, 12-s-trans conformation in rhodopsin by a bicycle-pedal isomerisation. The model reproduces the main experimental observations and explains how the protein makes the photoisomerisation path unique.

In the first step of the process of vision the rhodopsin molecule (a complex of 11-*cis* retinal and the protein opsin) absorbs a photon and undergoes a unique transition to the prelumirhodopsin intermediate¹. (For a review of recent work see refs. 2 and 3.) Some of the most definite information about this reaction emerged from picosecond laser⁴ and resonance Raman^{5,6} experiments as well as from studies of artificial pigments⁷⁻⁹. The transition from rhodopsin to prelumirhodopsin takes less than 6 ps (6×10^{-12} s) and the retinal moiety in this intermediate is characterised by strong resonance Raman lines that do not appear in rhodopsin⁵. The interaction with opsin directs the pathways of photoisomerisation of quite different retinal isomers to the same all-*trans* final form⁸. In spite of the potential usefulness of this information there is no unique interpretation for any of these findings and the nature of the first step remains unresolved. It is reasonable to assume that the rhodopsin molecule uses the light energy for the photoisomerisation reaction of 11-*cis* retinal as fast as possible or else the excited state will be deactivated by radiation and radiationless processes. It is not clear, however, how the photoisomerisation can be accomplished in as large a molecule as retinal in such a short time^{4,5} and how it can be done at all in a restricted active site¹⁰. Furthermore, there is no clue to the most intriguing mechanism by which the protein leads to only one final isomer from all the different initial complexes⁸.

Recent theoretical developments¹¹⁻¹⁵ make it possible to analyse the available experimental information and to calculate reliably ground and excited state equilibrium geometries^{12,13}, absorption spectra^{12,13} and resonance Raman spectra of retinal and related molecules (ref. 14 and my unpublished work with M. Karplus). It is also possible to simulate the dynamics of the photoisomerisation process and to monitor the conformational changes as a function of time¹⁵. In addition the interaction between the chromophore and hypothetical charges on the protein can also be evaluated¹⁶. The available theoretical approaches can be used to study various models for the visual excitation.

Here I propose a detailed model for the molecular events after the absorption of a photon by rhodopsin. The recently developed theoretical methods¹²⁻¹⁶ are used to test the model against the principal experimental facts about the first step of

the vision process and to stimulate the molecular dynamics of retinal during photoisomerisation. The following general assumptions are introduced in this model.

(1) The retinal is bound to a restrictive active site in which the cyclohexene ring, at one end of the molecule, is trapped in a hydrophobic cleft (this suggestion is supported by binding experiments⁹ and by analogy to other enzymes) and the other end of the molecule (the "tail end") is bonded by a Schiff base bond (C=N) to a lysine residue of the protein. (2) No major conformational changes of the protein are involved in the first step (for discussion see refs 2 and 3). (3) The light energy must be used for isomerisation around the *cis* double bond, since this is the main energetic barrier of the visual process and it would be rather difficult to perform fast thermal isomerisation after the dissipation of the light energy. (4) In addition to these assumptions which are the working hypotheses of the model it is also assumed that the chromophore in rhodopsin is set initially to the 11-*cis*, 12-*s-trans* conformation (this assumption is supported by ref. 7 as well as by the work reported here) and that the Schiff base nitrogen is protonated^{5,6}. It is further assumed that the protonated Schiff base bond is protected from hydrolysis by a negative group of the protein (a salt bridge) which increases the basicity of the Schiff base nitrogen.

Main features of the model

In the proposed model, described schematically in Figs 1 and 2, the chromophore is placed in the protein active site which is defined by assumptions (1) and (4). After absorption of light the *cis* bond is propagated to the tail end of the retinal chain by a 'bicycle-pedal'-like motion (see below). In this motion the molecule first changes its conformation by moving downhill on the potential surface of the excited singlet state and then crosses to the ground state once the C11-C12 torsional angle is approximately 90°. In the ground state the molecule becomes trapped in a strained all-*trans* (Fig. 2d) form, where the bond to the lysine residue forces the chain to bend and the salt bridge is stretched relative to its equilibrium distance in rhodopsin. This strained all-*trans* is the prelumirhodopsin intermediate in which the Schiff base nitrogen is less basic than in rhodopsin since its salt-bridge length is larger.

An important feature of the model is that all the isomerisation modes of different possible pigments can be executed by a concerted rotation about parallel pairs of double bonds ('bicycle-pedal' motion'), as described in Fig. 2. In particular, the *cis* conformation can propagate from different double bonds to the terminal C=N bond. Thus, the protein makes the retinal change its conformation along a one-dimensional reaction path (essential for the large photon efficiency) simply by constraining the ends of the molecule so that it must move in a 'bicycle-pedal' motion rather than in all the possible modes of motion which are possible in solution⁸.

Detailed examination of potential surface

The retinal intramolecular energy surface and its analytical derivatives were evaluated by the method already developed^{12,13}. Since the barriers for rotation around double bonds are of special importance here, we recalculated the potential parameters with particular emphasis on the reproduction of the

*Present address: Department of Chemistry, University of Southern California, Los Angeles, California 90007

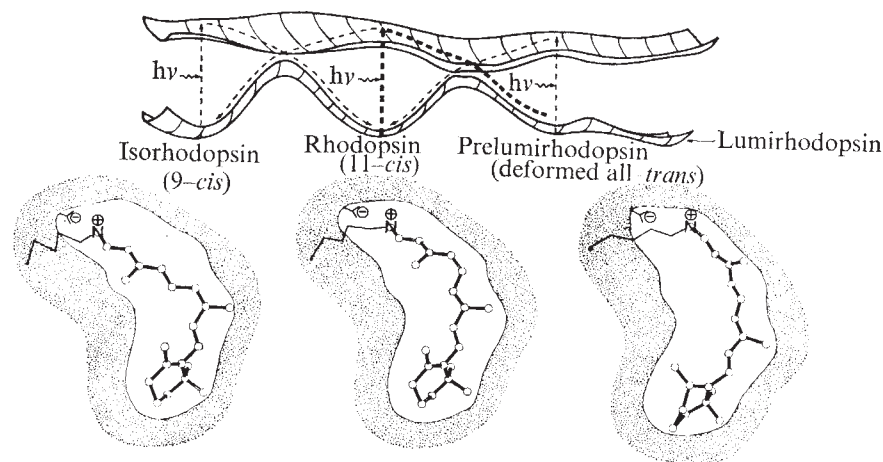


Fig. 1 A schematic description of the proposed model for the first step in the vision process. The figure presents the variation of the ground and excited state potential surfaces as a function of the adiabatic reaction coordinate and an effective perpendicular coordinate which represents all other coordinates. The change of the chromophore conformation in the enzyme active site are presented schematically below the surfaces. The thick broken line describes the primary visual process when the retinal chromophore of rhodopsin is excited by light to the first excited singlet, slides down, crosses to the ground state and becomes trapped in the Prelumirhodopsin conformation. The thin broken line describes the mechanism of reversible transition between different intermediates.

potential surfaces of the ground and the first excited singlet and triplet states of ethylene. The calculation includes a configuration interaction treatment of all the single excitations and the low lying double excitations. The parameters for the Schiff base nitrogen were determined by an extensive fit to many independent properties of nitrogen-containing compounds (my unpublished work with A. Lippicesela)¹². The σ electrons' charge distribution in the protonated Schiff base was determined by an all-valence electron approach¹⁶. In conventional semi-empirical quantum mechanical calculations the energy of the first excited singlet does not decrease smoothly as the C11–C12 double bond is twisted from 0° to 90° giving an activation barrier so that the singlet photoisomerisation becomes quite inefficient. We have found, however, that whenever the calculations include double configuration interaction treatment and Lowdin overlap correction^{12,13} the energy decreases monotonically on twisting to 90° without any activation barrier.

The effect of the active site was included in the calculation by taking into account the main features of steric restriction and electrostatic interaction. The steric restriction on both the ring and the tail was simulated by forcing the ring carbons and the α carbon of the lysine residue to stay close to their positions in 11-*cis*, 12-*s-trans* retinal (the starting conformation). The electrostatic interaction with the enzyme was simulated by the recently developed "microscopic" dielectric model¹⁶, using the atomic polarisability and average density of a typical protein (in this case lysozyme). It was found that while the effect of negative charge at various points along the chain can account for the spectral shifts between different pigments, its effect in modifying the surfaces is much smaller than that which could be expected in view of the sudden redistribution of the charges of the isolated chromophore on twisting of the C11–C12 bond to 90° (ref. 23). This effect is discussed below. Thus although the

presence of external charges can reduce the spacing between the ground and excited states in the 90° region, the molecular dynamics calculation was performed with only a single negative charge near the N⁶ of the lysine residue.

Simulation of photoisomerisation dynamics

If the transition between rhodopsin and Prelumirhodopsin involves the proposed *cis-trans* photoisomerisation, then any simulation of this process must reproduce the fact that the transition is faster than 6 ps. The simulation should also account for the large quantum efficiency of the reaction and for the concept of a complete isomerisation in a restricted active site. Here the "semiclassical trajectory approach" is used to investigate surface crossing in photoisomerisation¹⁵. In this approach one evaluates classically the intramolecular motion (trajectory) on the excited state surface and then uses the calculated variation of the coordinates with time to evaluate quantum mechanically the transition probability to the ground state. At each point of large transition probability the trajectory is split into two parts—one crosses to the ground state and the second continues on the excited state surface. Each of these trajectories is weighted by the corresponding transition probability. Here we are dealing with a strong singlet–singlet coupling and the adiabatic representation^{17,18} rather than the diabatic¹⁵ one is used. The result of a typical trajectory is presented in Figs. 3 and 4, which show that the molecule can perform isomerisation inside a restrictive active site in about 2×10^{-13} s. This fast isomerisation is due to the small moment of inertia of the bicycle-pedal motion. Figures 3 and 4 also show that the molecules cross to the ground state with 30% probability during a single pass through the 90° region. Thus after about three passes (which corresponds to about 6×10^{-13} s) the probability of the molecule being in the ground state is almost 100%. The calculation clearly shows that a complete

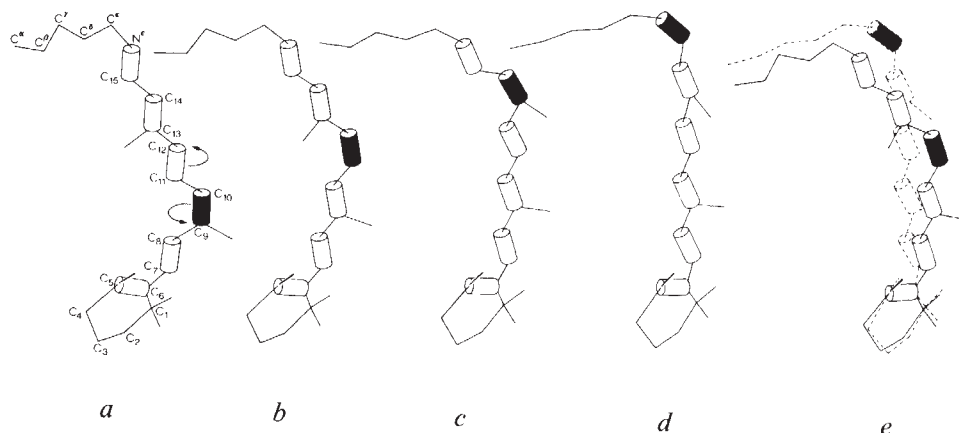
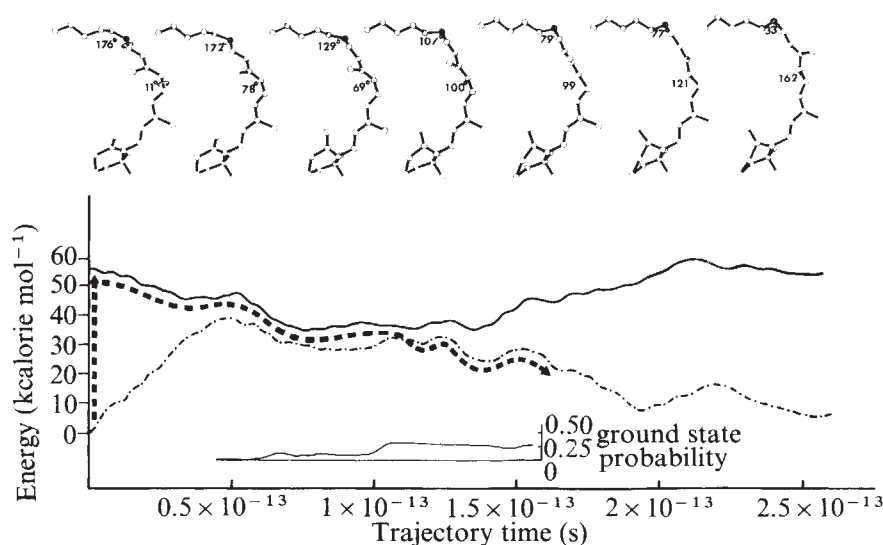


Fig. 2 Typical elements of the 'bicycle-pedal' mechanism. *a-d*, 9-*cis*, 11-*cis*, 13-*cis* and 15-*cis* isomers, respectively; *e*, the one-step concerted rotation about the 11–12 and 15–16 bonds which is the proposed motion in the visual excitation. Clearly with a system of parallel double bonds the *cis* bond can be transferred to any point without large shift into the atomic coordinates.

Fig. 3 Computer simulation of the first step in the vision process. The figure presents a single trajectory which begins on the first excited singlet and crosses to the ground state at the point of maximum transition probability ($\phi_{11-12} = 95^\circ$). The potential energy of the ground and the first excited singlet are described by the broken and solid lines, respectively, as a function of the trajectory time while the conformational changes along the trajectory are presented above with the corresponding values of ϕ_{11-12} and ϕ_{15-16} . The trajectory was obtained in the complete Cartesian space starting with the ground state equilibrium coordinates and zero kinetic energy. The ring and the tail ends were constrained by assigning masses of 1,000 atomic units to the C1 and C5 ring carbons and to the C α of the lysine residue. The humps on the potential surface along the trajectory are due to transfer of kinetic energy to potential energy in several bond-stretching, angle-bending and torsion vibrations. The trajectory describes a dynamical rather than an equilibrium process and the dissipation of the excess torsional potential energy in the many normal modes (~ 200) of the system is not complete. Thus in some energy fluctuations much of the kinetic energy is converted back to potential energy. The probability of being at the ground state $a_0(\tau)$ is given as a function of the trajectory time as:



$$a_0(\tau) = \int_0^\tau \left[\langle \Psi_0 | \frac{\partial \Psi_1}{\partial t} \rangle a_1(t) \exp \left\{ -(i/\hbar) \int_0^t (E_1 - E_0) dt' \right\} \right] dt$$

where Ψ_0 and Ψ_1 are the ground and excited state electronic wave functions and E_0 and E_1 are the corresponding potential energies. In the actual evaluation of the *cis-trans* quantum yield the trajectory is split first at $\phi_{11-12} = 82^\circ$ where a trajectory with ~ 0.1 probability crosses to the *cis* ground state while the main trajectory continues in the excited state up to the point $\phi_{11-12} = 95^\circ$ where it splits again with ~ 0.2 crossing to the ground state (this trajectory is shown in Fig. 4), while another trajectory proceeds on the excited state with 0.7 probability.

This splitting procedure continues until the probability of the excited state trajectory is less than 0.1.

cis-trans photoisomerisation can be accomplished in less than 6 ps in contrast to previous suggestions⁴. Another important point is the fact that the calculated quantum efficiency is larger than 50%, indicating that the model system is almost as efficient as rhodopsin.

In contrast to the 'bicycle-pedal' motion the conventional photoisomerisation path involves a large rotation about the *cis* bond of either the ring¹⁰ or the tail end. The resulting motion will be very slow due to the large moment of inertia and high probability for collision with the walls of the active site.

Resonance Raman and absorption spectra

The resonance Raman (RR) technique can provide unique information about the visual pigment as it measures the vibrational spectrum of the retinal inside the active sites of the protein. Recent measurements have shown that prelumirhodopsin is characterised by lines in the 850–950 cm^{-1} region⁸, which are not observed in rhodopsin or in Schiff base of retinal in solution. Our method for evaluation of vibronic transition intensities¹¹ can be used for direct calculation of the RR spectrum. This approach reproduces the main features of the RR spectra of Schiff bases of retinal in solution¹⁴. Model calculations have shown that the special lines of prelumirhodopsin can be reproduced if the chromophore is strained (my unpublished results with M. Karplus). The calculation of the RR spectrum of the present model of prelumirhodopsin (Fig. 2d) also reproduces two very strong lines at the 900 cm^{-1} region. These are due to a relatively large change in the geometry of the strained chromophore on electronic excitation.

Previous calculations of the dependence of the absorption maximum (λ_{max}) on external charges are reviewed in refs 2 and 3. In contrast to these works I include in my calculations the major effect of the protein polarisability¹⁶. This enables me to analyse properly the effect of external charges on λ_{max} . The degree of protonation was explicitly taken into account by calculating the σ charges of the retinal, the proton and the counter ion by an all-valence electron approach. I found that the redshift between rhodopsin (498 nm) and a protonated Schiff base of retinal in solutions (440–480 nm) can be accounted

for by a calculated redshift of λ_{max} from 460 to 485 nm which is obtained when the salt bridge is placed in a more polar environment than the rest of the chromophore. An additional redshift of about 30 nm is obtained when the salt bridge is stretched from 2.5 to 4.5 Å. A further redshift of as much as 40 nm is obtained by another charged group properly placed near the chain, whereas a redshift of about 20 nm is obtained by $\sim 10^\circ$ twist about a double bond (obviously the shift can be increased by a larger twisting). Thus a significant part of the redshift between prelumirhodopsin (543 nm) and rhodopsin can be accounted for by a stretching of the salt bridge. The possibility, however, cannot be excluded that the redshift is due to the effect of double bonds twisting and/or changes in the distances between the chain and some negative or positive group. The inclusion of the polarisability reduces by about half the previously assumed effect of external charges^{2,3}. Thus, it is likely that some part of the redshift between prelumirhodopsin and rhodopsin is due to twisting of double bonds. Essentially, Figs 1 and 3 show that a redshift of any magnitude can be obtained depending on the equilibrium value of ϕ_{11-12} , which is determined by the balance between the intramolecular potential of the chromophore and the constraint of the protein.

Another related fact, which introduces difficulties in other models¹⁹, is the finding that the transition dipole is almost parallel in rhodopsin and the different bleaching intermediates²⁵. This is clearly reproduced by my model which keeps all the double bonds parallel during the different stages of visual excitation.

Quantum efficiency and reversibility

An interesting and so far unresolved problem concerns the facts that prelumirhodopsin is a common intermediate for both rhodopsin and isorhodopsin I (9-*cis*)^{5,10} and that the quantum efficiency of rhodopsin is about 0.7 while that of both isorhodopsin I and isorhodopsin II (9,13-*dicis*) is about 0.3 (ref. 8). This behaviour can also be studied by the semiclassical trajectory approach¹⁵. The calculated quantum efficiency for my rhodopsin model is about 0.6. A preliminary calculation starting with isorhodopsin I gave about 0.3 probability for

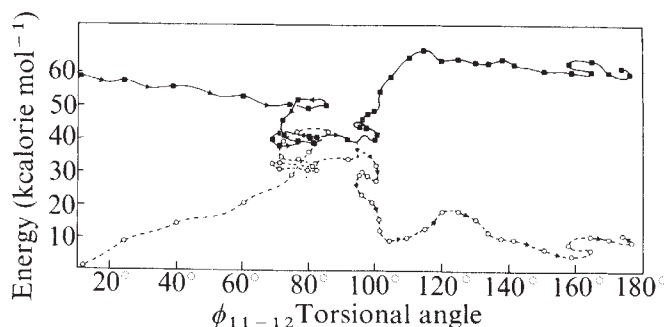


Fig. 4 The relationship between the potential surfaces and the ϕ_{11-12} torsional angle along the sample trajectory which is presented in Fig. 3. The trajectory time intervals (6×10^{-4} s) are represented by (○) and (●) for the ground and excited state, respectively. The actual potential surface taken by the trajectory is designated by (—). The figure shows that the molecule forms some 'activated complex' and stays a relatively long time in the 90° region. The different values of the potential energy for similar values of ϕ_{11-12} are due to changes in other coordinates. Thus after the surface crossing ϕ_{11-12} varies slowly and the changes in the potential are due to changes in ϕ_{13-16} and ϕ_{9-10} as well as in various bond lengths and bond angles.

photoisomerisation in a 'bicycle-pedal' motion to rhodopsin (a similar part proceeds directly to prelumirhodopsin in a concerted rotation about the 9–10 and 15–16 bonds). With a second photon the rhodopsin proceeds with 0.7 probability to prelumirhodopsin. This suggestion can account for the fact that isorhodopsin I is about half as efficient as rhodopsin. The isomerisation of the isorhodopsin II (9,13-*dicis*) can be explained as described in Fig. 5: the molecule can move directly to prelumirhodopsin by a concerted bicycle-pedal rotation about the 9–10 and 13–14 bonds. Alternatively the molecule can change its conformation to 11,13-*dicis* and then directly to all-*trans* by a second 'bicycle-pedal' motion of the two adjacent *cis* bonds.

The reversibility of photoisomerisation between prelumirhodopsin, rhodopsin and isorhodopsin (my unpublished results with M. Karplus) is accounted for by the fact that absorption of photons in one isomer populates its excited state, which can be relaxed to the ground states of the other isomers (Fig. 1).

Alternative explanations

Although other models of the visual excitation have been proposed^{1,10,19-24}, they did not provide unique isomerisation paths nor a detailed proposal for the conformation of prelumirhodopsin. Nevertheless, these models have some interesting features which are worth considering in terms of alternative explanations for the observations.

Recent calculations²³ have shown that the excited state charge distribution of the chromophore might change drastically on twisting of the C11–12 bond to 90°. A "short-lived electrical signal" mechanism was proposed as the primary event in the vision process. Such a mechanism seems to be unlikely as long as it implies a very short pulse rather than an irreversible change. The strong dependence of the excited state polarisation on the C11–12 torsional angle, however, presents the intriguing possibility that prelumirhodopsin has a twisted double bond where an external charge stabilises the excited singlet so that it becomes a ground state. If such a mechanism were operative along the 'bicycle-pedal' pathway it could account for many observations. My calculation, however, never reproduced this stabilisation effect, even with drastic changes of the semi-empirical parameters: that is, if the retinal charge distribution is not allowed to change in response to any external charge effect, then an optimally placed negative group could push the twisted excited state as much as 50 kcalorie mol⁻¹ below the twisted ground state. When, however, a complete charge redistribution was allowed in my calculations, I was unable to push the excited state more than 5 kcalorie mol⁻¹ below the ground state. This is because the presence of the negative

charge near various points with a positive charge in the excited state of the isolated molecule induces positive charges in the ground state at the same points, which in turn makes the excited state charge less positive. I feel, however, that my model calculations cannot exclude completely the possibility that the chromophore in prelumirhodopsin is trapped in a twisted form by a strong interaction with some charged group of the protein.

Another suggestion is that the protonated Schiff base releases a proton in the primary event in the vision process²⁴. The problem with this model is that the photoisomerisation must then be accomplished by thermal energy (this would cost about 25 kcalorie mol⁻¹) and the light energy will be wasted. Furthermore, the proton dissociation will cause a blueshift rather than a redshift in the absorption spectrum of prelumirhodopsin.

The possibility that the photoisomerisation is accomplished through the triplet state rather than the singlet state cannot be excluded completely. Recent experiments²⁶, however, support the singlet mechanism, and my study has shown that there is a significant barrier for isomerisation in the triplet. An estimate of the triplet-singlet coupling indicates that this coupling gives less efficient surface crossing than the non-adiabatic singlet-singlet coupling.

We cannot disregard completely the possibility that the prelumirhodopsin intermediate is not in a strained all-*trans* conformation but in other conformations along the reaction path between 9-*cis* and 11-*cis*. Such a ground state conformation which requires a large twist about the 10–11 single bond as well as about the 9–10 and 11–12 double bonds might be possible if rhodopsin actually binds retinal in the 11-*cis*, 12-*s-cis*-form rather than 11-*cis*, 12-*s-trans* form. The photon efficiency of this system, however, should be less than 0.3 since the sum of the probabilities of isomerisation to 9-*cis* and of return to 11-*cis* is larger than 0.7.

It is clearly possible that the chromophore in prelumirhodopsin is in a much more twisted form than the one obtained by my potential surfaces. That is, the protein can provide such an external constraint which will stop the isomerisation in an earlier stage where ϕ_{11-12} and other torsional angles are strongly twisted. Obviously in this case the strain energy at prelumirhodopsin will be higher than that implied by the present calculation.

Implications of the model

The model presented here accounts for the main observations concerning the first step in the vision process. It is consistent with the ultra-fast transition from rhodopsin to prelumirhodopsin; the changes in the RR and absorption spectra during the transition, and the conservation of the direction of the transition moment. The model explains the uniqueness of the photoisomerisation process and its high quantum efficiency as well as the reversibility of transitions between the different pigments. The model is also consistent with present understanding of steric effects and charge stabilisation: that is, a

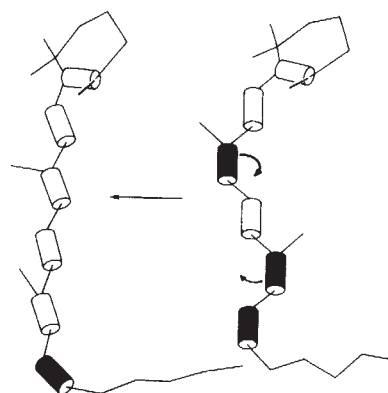


Fig. 5 A possible bicycle-pedal path for isorhodopsin II (9,13-*dicis*). The molecule moves in a concerted rotation about the 9–10 and 13–14 bonds directly to prelumirhodopsin.

steric strain by the protein is not effective if it can be relaxed by small displacements of the corresponding atoms. On the other hand, the steric effect may be important when the molecule is trapped in a twisted form by a constraint which can be relaxed only by large conformational changes. The charge effect (which was proposed as a very efficient way of controlling enzymatic reactions¹⁶) is used here with a salt bridge to the protonated Schiff base, whereas a fine tuning of the system can be obtained by other charged residues optimally placed along the chain.

It now seems that the first step of the vision process can be described as follows. Light absorption by rhodopsin populates the first excited electronic singlet state of the 11-*cis* retinal chromophore (which is bound at both ends as proposed here). The excited retinal slides in a very fast 'bicycle-pedal' motion to the 90° twisted singlet and then crosses to the ground electronic state while continuing to slide towards the *trans* form (Fig. 1). The protein defines a unique isomerisation valley for the conjugated chain by restricting the motion of both the ring end and the tail end and by some charge stabilisation. The use of the protonated Schiff base system with a proper charge stabilisation guarantees efficient singlet-singlet crossing and allows the system to perform very efficient isomerisation before other deactivation processes can take place. After the surface crossing step, the chromophore is trapped in a high energy strained form (prelumi rhodopsin) and the basicity of the Schiff base nitrogen is reduced as a result of the increase in distance between the negative residue and the N⁶ relative to that in rhodopsin. In the next stage of the reaction (lumirhodopsin) λ_{max} is blueshifted by ~45 nm relative to prelumi rhodopsin. This shift might indicate changes in the polar environment of the Schiff base bond and/or other parts of the chromophore, as well as a decrease in the torsional deformation of some double bonds of the chromophore. Such changes are possible (when the temperature is increased and prelumi rhodopsin is changed to lumirhodopsin) as a result of a partial release of the chromophore strain which involves some small conformational changes of the protein. The release of strain and the new balance of the electrostatic interactions lead to more significant conformational changes of the protein during the steps lumirhodopsin to meta I to meta II. The spectral changes in the meta I to meta II step¹ indicate that the Schiff base proton is dissociated after further decrease in the basicity of the Schiff base nitrogen. In contrast to the situation in rhodopsin the non-protonated Schiff base in meta II can undergo hydrolysis by water. The protein conformational changes serve to convert the light signal into an electrical membrane current, somewhere along the pathway between prelumi rhodopsin and meta II. These

stages, however, do not involve further isomerisation of the chromophore.

It seems clear that in designing this photoconverter machine the two boundary conditions, which in turn establish the 'bicycle-pedal' mechanism, are important. In this way the system is forced to move in what is effectively one dimension along the complex energy surface as the *cis* bond is transferred backwards and forwards. Thus it seems that a promising way to simulate the surface crossing in rhodopsin might be achieved by a chemical approach of building a model compound with a Schiff base bridge between the ring to the end of the chain.

Yoshizawa and Wald have proposed²⁷ that the retinal molecule in this intermediate is in a strained all-*trans* conformation. Although the details of the transition between rhodopsin to prelumi rhodopsin and the binding of the cyclohexene ring¹ were not clear at that stage, some features are common to both models.

I thank M. Levitt for encouragement and critical discussion and M. Karplus and B. Honig for stimulating discussions in the early stages of this work. This work was supported in part by the United States-Israel Binational Science Foundation and by the European Molecular Biology Organization.

Received January 7; accepted March 3, 1976.

1. Wald, G., *Science*, **162**, 230-239 (1968).
2. Honig, B., and Ebrey, T. G., *A. Rev. Biophys. Bioengng.*, **3**, 151-177 (1974).
3. Ebrey, T. G., and Honig, B., *Q. Rev. Biophys.*, **8**, 130-184 (1975).
4. Busch, G. E., Applebury, M. L., Lamola, A. A., and Rentzepis, P. M. R., *Proc. natn. Acad. Sci. U.S.A.*, **69**, 2802-2805 (1972).
5. Oseroff, A. R., and Callendar, R. H., *Biochemistry*, **13**, 4243-4248 (1974).
6. Lewis, A., and Spoonhower, J., in *Neutron, X-ray and Laser Spectroscopy in Biophysics and Chemistry*, 342-376 (Academic, New York, 1974).
7. Chan, W. K., Nakanishi, K., Ebrey, T. G., and Honig, B., *J. Am. chem. Soc.*, **96**, 3642-3644 (1974).
8. Crouch, R., Purvin, V., Nakanishi, K., and Ebrey, T. G., *Proc. natn. Acad. Sci. U.S.A.*, **72**, 1538-1542 (1975).
9. Blatz, P. E., Lin, M., Balasubramanian, P., Balasubramanian, V., and Bewhurst, P. B., *J. Am. chem. Soc.*, **91**, 5930-5931 (1969).
10. Kropf, A., in *Processing of Optical Data by Organisms and by Machines*, Course 43, Proc. Int. School Phys. Enrico Fermi (edit. by Reichadi, W.), 28-43 (Academic, New York, 1969).
11. Honig, B., Warshel, A., and Karplus, M., *Acc. chem. Res.*, **8**, 92-100 (1975).
12. Warshel, A., and Karplus, M., *J. Am. chem. Soc.*, **94**, 5612-5625 (1972).
13. Warshel, A., and Levitt, M., *Quantum Chemistry Program Exchange*, Program No 247 (University of Indiana, Bloomington, 1973).
14. Warshel, A., and Karplus, M., *J. Am. chem. Soc.*, **96**, 5577-5689 (1974).
15. Warshel, A., and Karplus, M., *J. Chem. Phys. Lett.*, **32**, 11-17 (1975).
16. Warshel, A., and Levitt, M., *J. molec. Biol.* (in the press).
17. Tully, J. C., and Preston, A. K., *J. chem. Phys.*, **55**, 562-572 (1971).
18. Miller, W. H., and George, T. F., *J. chem. Phys.*, **56**, 5637-5652 (1972).
19. Abrahamson, E. W., and Ostroy, S. E., *Prog. Biophys. molec. Biol.*, **11**, 179-215 (1967).
20. Honig, B., and Karplus, M., *Nature*, **229**, 558-560 (1971).
21. Kakitani, T., and Kakitani, H., *J. Phys. Soc. Japan*, **38**, 1455-1463 (1975).
22. Yoshizawa, T., and Wald, G., *Nature*, **214**, 566-571 (1967).
23. Salem, L., and Bruckmann, P., *Nature*, **258**, 526-528 (1975).
24. Thomson, A. J., *Nature*, **254**, 178-179 (1975).
25. Wright, W., Brown, P., and Wald, G., *J. gen. Physiol.*, **62**, 509 (1973).
26. Aichalal, A., Honig, B., Ottolenghi, M., and Rosenfeld, T., *J. Am. chem. Soc.*, **97**, 2161-2166 (1975).
27. Yoshizawa, T., and Wald, G., *Nature*, **197**, 1279-1286 (1963).

The nucleus of Centaurus A

A. C. Fabian, D. Maccagni*, M. J. Rees & W. R. Stoeger

Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA, UK

The nucleus of Centaurus A contains a very compact radio component, and has recently been discovered to be a variable X-ray source. We argue that these phenomena arise from accretion on to a massive black hole. Possible implications for galactic nuclei in general are briefly discussed.

CENTAURUS A, at a distance of ~5 Mpc, is the closest radio galaxy. Its total radio power is now only ~10⁴² erg s⁻¹. The energy content of the very extended radio lobes is, however, ~10⁶⁰ erg, and this suggests that Cen A may once have been a powerful radio source of the same type as, for instance, Cygnus

A. Recent observations of apparent radio and X-ray variability provide evidence for continuing activity in the nucleus of Cen A. Even though this activity involves a much smaller power output than occurs in QSOs and strong radio sources, it may suggest clues to the nature of active galactic nuclei in general. We discuss here the implications of the new data (particularly the X-ray variability), and argue that they perhaps indicate the presence of a massive (~10⁷M_☉) black hole in Cen A.

Kellermann¹ has detected a central radio component with luminosity ~4 × 10⁴⁰ erg s⁻¹ whose spectrum cuts off below ~30 GHz. If the cutoff results from synchrotron self absorption, the size of this component would be only ~10¹⁶ cm. There is some evidence^{1,2} for variability on time scales as short as a day.

Cen A is also an X-ray source. Davison *et al.*³ found that the X-ray intensity had varied over a two-year interval, and therefore

*Permanent address: Laboratorio di Fisica Cosmica & Tecnologie Relative, CNR, Milano, Italy.

proposed that it was associated with a very compact nuclear source. This variability has been confirmed by independent measurements⁴⁻⁶; indeed Winkler and White⁴ reported X-ray variability of a factor 1.8 ± 0.2 within only 6 d. The spectrum is hard, and there is evidence for variability at photon energies up to $\gtrsim 100$ keV (ref. 7). The X-ray power at < 100 keV is $\sim 5 \times 10^{43}$ erg s⁻¹. Hall *et al.*⁷ report a γ -ray continuum extending up to ~ 5 MeV, with possible lines at 1.6 MeV and 4.5 MeV; and very high energy ($\gtrsim 300$ GeV) γ rays have been claimed⁸.

Previous discussions of the nature of the X-ray source in Cen A have concentrated on non-thermal models⁹⁻¹¹, and do not specify the source of the necessary relativistic electrons. The idea that massive black holes may lurk in galactic nuclei^{12,13} is appealing for several reasons. A black hole seems the likely endpoint for the evolution of a supermassive star or dense star cluster (these being two alternative models for an active galactic nucleus) and accretion on to a black hole can itself account for various phenomena associated with galactic nuclei. If it (or its progenitor) had generated the entire energy content of the extended radio source, the black hole in Cen A would be $\gtrsim 10^7 M_\odot$. In what follows we consider a hole of mass $10^7 M_\odot$ solar masses (and Schwarzschild radius $r_s \simeq 3 \times 10^{12}$ M_7 cm).

Accretion on to massive black holes

If the X rays are thermal bremsstrahlung, then the hard spectrum implies that the temperature is $T \simeq 10^9$ K (in reality, of course, one might expect different parts of the source to span a range of temperature). If the emission comes from gas occupying a sphere of radius r around the black hole, with a volume 'filling factor' f , then the approximate mean particle density n must be

$$n \simeq 10^{14} \left(\frac{r}{r_s} \right)^{-3/2} f^{-1/2} M_7^{3/2} \text{ cm}^{-3} \quad (1)$$

The ratio of the bremsstrahlung cooling time

$$t_{\text{br}} \simeq 3 \times 10^{11} T^{1/2} n^{-1} \text{ s}$$

to the free fall time scale

$$t_{\text{ff}} \simeq 10^3 M_7 \left(\frac{r}{r_s} \right)^{3/2} \text{ s}$$

is then

$$t_{\text{br}}/t_{\text{ff}} \simeq f^{1/2} M_7^{1/2} \left(\frac{T}{10^9 \text{ K}} \right)^{1/2} \quad (2)$$

For the parameters appropriate to Cen A, this ratio never greatly exceeds unity (and it can be made $\ll 1$ by taking a small filling factor); so the efficiencies are adequate provided that the gas can be heated to $T \gtrsim 10^9$ K.

If ϵ denotes the efficiency of the accretion process in converting rest mass into X rays, then the required mass inflow rate is $\sim 10^{-3} \epsilon^{-1} M_\odot \text{ yr}^{-1}$. The corresponding particle density, at distance r from the hole, would be

$$n(r) \simeq 3 \times 10^{11} \epsilon^{-1} f^{-1} \left(\frac{r}{r_s} \right)^{-3/2} M_7^{-2} \left(\frac{v_r}{v_{\text{ff}}} \right)^{-1} \text{ cm}^{-3} \quad (3)$$

If transverse motions or pressure gradients slow down the inward velocity, so that the radial velocity v_r becomes $\ll v_{\text{ff}}$, note that n is correspondingly higher for a given mass accretion rate. Equations (1) and (3) cannot be compared directly, since

v_r , f and T are probably all functions of radial distance r . They do indicate however that even radially infalling material may produce the observed X rays if $\epsilon \sim 10^{-3}$, $f \sim 10^{-1}$ and $M_7 \sim 1$.

When the specific angular momentum of the captured material greatly exceeds $a_{\text{crit}} = 2cr_s$, we expect the formation of a significant accretion disk extending well beyond $6r_s$; and it is well known that in this situation a high radiative efficiency $\epsilon \simeq 0.1$ is almost guaranteed. The 'standard' model¹⁴, however, then predicts disk temperatures of only $T \lesssim 10^5$ K for $M_7 \gtrsim 1$ (no X rays at all). This is because the required luminosity is far below the 'critical' luminosity $\sim 1.3 \times 10^{45} M_7 \text{ erg s}^{-1}$ for such a massive black hole. The inner, radiation pressure supported, region of the disk is therefore thin. The factors f and v_r/v_{ff} in equation (3) are then both very small, and n becomes so high that bremsstrahlung is efficient even at quite low temperatures.

The other well studied type of accretion is that where the inflow is radial (zero angular momentum) and laminar. We see from equation (2) that $t_{\text{br}}/t_{\text{ff}}$ increases with T . The problem here is therefore that if this factor is to be $\lesssim 1$ near the hole, it would have been $\ll 1$ at larger radii. The gas would therefore have remained cool, and the radiative output derived from 'pdV work' would be low¹⁵. (We avoid introducing the usual concept of 'accretion radius' in this context, because, (1) at large distances the gravitational field of the whole galactic nucleus dominates that of the hole itself, and (2) most of the gas may in fact be produced at a radius where the escape velocity is supersonic.)

Quasi-spherical accretion

Of course there are many ways in which the above models could be modified to account for reasonably efficient production of hard X rays. For instance, realistic disks may involve hot coronae, magnetic flares, and so on. Alternatively, as emphasised by Mészáros¹⁶, continual dissipative heating from turbulence or magnetic field reconnection could raise the efficiency and temperature in the radial inflow picture; but, where cooling is efficient, this model is still constrained by the requirement that

$$kT(r) \lesssim 10^{12} \left(\frac{r}{r_s} \right)^{-1} \frac{t_{\text{br}}}{t_{\text{ff}}}$$

We wish to point out, however, that the X-ray data are most readily explained by an 'intermediate' model where the infalling material has a small but significant specific angular momentum a in the range (say) $1-30 \times a_{\text{crit}}$. A typical element of gas must then lose its excess angular momentum before it can fall into the hole. (This type of process can in principle yield higher efficiencies than disk accretion. Consider, for instance, two identical (and very small) gas clouds dropping towards a Schwarzschild black hole in orbits having equal opposite angular momentum. If these clouds collided at the point r_{min} when their orbits pass closest to the hole they would come momentarily to rest. If the energy of impact were radiated away, the efficiency would be $1 - [1 - (r_s/r)]^4$. For clouds with the critical angular momentum, a_{crit} , the corresponding volume of r_{min} is $2r_s$. This yields an efficiency of 29% (neglecting a correction for the fraction of radiation that would itself be swallowed).)

We can envisage two possibilities: firstly, the accreted material may all have the same sense of rotation. In this case a disk will initially form at a radius $\sim (\langle a \rangle / a_{\text{crit}})^2 r_s$. If the flow continues steadily for a long time, then viscous effects in the disk may cause its rim to grow outwards, but (except in the plane of the disk) inflowing gas would still join the disk after passing through a shock at radius $\sim (\langle a \rangle / a_{\text{crit}})^2 r_s$. Secondly, even if there were no steady systematic rotation, a unit mass of infalling gas might have some angular momentum $\sim a$ arising from small random transverse velocities far from the hole. If $a > a_{\text{crit}}$, this material could not be swallowed until it had lost its angular momentum by colliding with counter-rotating fluid elements near the hole.

In either of the above cases, infalling material will be shocked at a radius $r_{\text{shock}} \simeq (\langle a \rangle / a_{\text{crit}})^2 r_s$, the temperatures attained being $\sim 10^{12} (r_s / r_{\text{shock}})$ K. Thus if $r_{\text{shock}} \lesssim 10^3 r_s$ the temperatures would be high enough to produce a bremsstrahlung spectrum extending up to photon energies of at least 100 keV. Moreover if $t_{\text{br}} \lesssim t_{\text{ff}}$ the shocked gas will have time to cool down before being swallowed, thus ensuring reasonable efficiency. The 'impulsive' form of heating from shocks can in this case result in much higher temperatures than from a continual dissipative release of energy.

Most of our discussion is independent of which of these two alternatives applies, and is therefore insensitive to the actual origin of the accreted gas. One interesting possibility is that the gas results from tidal breakup of stars passing close to the black hole¹³. This mechanism can plausibly provide an adequate supply of material; moreover this material would necessarily have angular momentum corresponding to orbits near the 'tidal radius' of the black hole— $\lesssim 10 a_{\text{crit}}$ for appropriate parameters¹³.

The variability and spectrum of the X rays would, in this scheme, be determined by the details of the inflow, which in turn depends on the origin of the accreted gas. If $M_7 \simeq 1$ –10 there is, however, no problem with accounting for variability on any time scale down to hours.

For the spectrum to extend into the MeV band, shocks would need to occur where $r/r_s \lesssim 100$. Note that there would then be additional cooling due to electron bremsstrahlung, pair production and 'Comptonisation'. These processes would reduce the cooling time, thus improving the efficiency and strengthening the need for impulsive rather than dissipative heating.

According to Hall *et al.*⁷, the spectrum extends up to ~ 5 MeV, and this permits us to set a tentative rough lower limit of $r \gtrsim 10^{15}$ cm to the source dimensions. This limit (which is somewhat geometry dependent) corresponds to the requirement that the 'column density' of hard X-ray photons in the source should not be so high that the γ rays produce e^-e^+ pairs before escaping¹⁷.

The radio source in the nucleus

The (probably variable) radio component, whose dimensions are also 10^{15} – 10^{16} cm, may well be associated with the region that emits the variable X rays. Indeed, the same shock waves which heat the infalling gas may also generate the relativistic electrons. This possibility cannot be explored quantitatively without estimating the magnetic field structure in the gas, or discussing acceleration mechanisms. One can, however, exploit an analogy with supernova remnants such as Cas A, where shock waves involving velocities somewhat lower than we envisage have certainly converted $\gtrsim 1\%$ of the available energy into relativistic electrons. The very high energy γ rays reported by Grindlay *et al.*⁸ are presumably somehow associated with the non-thermal radio source.

Radio variability on a time scale of months or years is a well known occurrence in many QSOs, and in the nuclei of some strong radio galaxies. It is commonly interpreted as a succession of 'outbursts', which each produce an expanding cloud of relativistic plasma. VLBI measurements provide evidence for changing structure in these components, and it has been widely argued^{18,19} that the outbursts represent successive supernova-style events each releasing $\sim 10^{53}$ erg: the inferred energy per outburst, the event rate ($\sim 1 \text{ yr}^{-1}$) and the mean luminosity of the galactic nucleus are all compatible with this view. Kellermann¹ emphasises that the nucleus of Cen A resembles a scaled-down version of these other variable sources. The event rate is, however, higher and the mean power output much lower. It seems implausible that supernovae are going off at a rate of 1 d^{-1} in Cen A, which is an order of magnitude less luminous than QSOs. Instead, it seems that the radio outbursts in active nuclei involve flares whose energy is correlated with the mean luminosity (as expected in an accretion picture), rather than supporting a

'building block' model where the energy per outburst is standardised. Moreover, we have shown that the massive black-hole hypothesis accounts naturally for the emission of variable thermal X rays with a hard spectrum. (That 3C279 produced optical flares involving $\gtrsim 10^{54}$ erg (ref. 20) raises problems for the multiple-supernova model in the opposite direction.)

If the nucleus of Cen A involves supernova-type events in a dense star cluster massive and long lived enough to have powered the extended radio source, then the present activity should be spread over a ~ 1 pc region. This corresponds to an angle $\sim 0.04''$ —large by the standards of VLBI techniques. Thus if the radio source, when it has been observed for a longer interval, fails to display any 'jitter', this would support the kind of model proposed here involving a massive black hole. It would not, of course, preclude the possibility that a dense star cluster is indirectly involved, as envisaged by Hills¹³.

Discussion

Accretion of material with low, but not negligible, specific angular momentum by a massive black hole can be an efficient means of producing hard X rays. Mészáros¹⁶ has emphasised that dissipative heating may greatly enhance the efficiency when $a < a_{\text{crit}}$, and we find that shocks involving velocity discontinuities of $\gtrsim 0.1c$ can occur in the intermediate case of $a_{\text{crit}} < a < 30 a_{\text{crit}}$. Radiation from behind such shocks can produce a hard spectrum extending up to $\gtrsim 1$ MeV. This mechanism can account for the observed X-ray emission from the nucleus of Cen A. The variability in the X-ray source may arise from changes in the mass flow rate, or amount and distribution of angular momentum in the infalling gas. (An analogous model may be relevant to Cygnus X-1 (refs 21–23).) 'Standard' accretion models^{14,15} fail to account for hard X-ray emission from massive black holes.

A similar model may also apply to the nucleus of 3C390.3, which is possibly a variable X-ray source²⁴, and such objects as NGC4151 and 3C273. The strong X-ray absorption observed²⁵ in these last two suggest again that the nucleus is responsible. An alternative scheme involving multiple stellar mass 'events' in a dense star cluster might be tested by searching for jitter with VLBI techniques. It is also possible in this case that the X-ray absorption might change if a substantial fraction of the absorbing material lies in the star cluster. If the analogy between Cen A and more powerful galactic nuclei is genuine, the implications for theories of extragalactic violent activity in general would be basic and fundamental.

D. M. acknowledges an Academia dei Lincei–Royal Society Fellowship.

Received February 12; accepted March 11, 1976.

- ¹ Kellermann, K. I., *Astrophys. J. Lett.*, **194**, L135–L137 (1974).
- ² Fogarty, W. G., and Schuch, N. J., *Nature*, **254**, 124–125 (1975).
- ³ Davison, P. J. N., Culhane, J. L., Mitchell, R. J., and Fabian, A. C., *Astrophys. J. Lett.*, **196**, L23–L25 (1975).
- ⁴ Winkler, P. F., and White, A. E., *Astrophys. J. Lett.*, **199**, L139–L142 (1975).
- ⁵ Grindlay, J. E., Schnopper, H., Schreier, E. J., Gursky, H., and Parsignault, D. R., *Astrophys. J. Lett.*, **201**, L133–L136 (1975).
- ⁶ Stark, J. P., Davison, P. J. N., and Culhane, J. L., *Mon. Not. R. astr. Soc.*, **174**, 35P–40P (1976).
- ⁷ Hall, R. D., *et al. Proc. 14th Int. Cosmic Ray Conf.*, **1**, 84–88 (1975).
- ⁸ Grindlay, J. E., Helmlen, H. F., Hanbury Brown, R., Davies, J., and Allen, L. R., *Astrophys. J. Lett.*, **197**, L9–L12 (1975).
- ⁹ Tucker, W., Kellogg, E., Gursky, H., Giacconi, R., and Tananbaum, H., *Astrophys. J.*, **180**, 715–724 (1973).
- ¹⁰ Perola, G. C., and Tarengi, M., *Astr. Astrophys.*, **25**, 461–465 (1973).
- ¹¹ Grindlay, J. E., *Astrophys. J.*, **199**, 49–53 (1975).
- ¹² Lynden-Bell, D., *Nature*, **233**, 690–694 (1969).
- ¹³ Hills, J. G., *Nature*, **254**, 295–298 (1975).
- ¹⁴ Shakura, N. I., and Sunyaev, R. A., *Astr. Astrophys.*, **24**, 337–355 (1973).
- ¹⁵ Pringle, J. E., Rees, M. J., and Pacholczyk, A. C. T., *Astr. Astrophys.*, **29**, 179–184 (1973), and references cited therein.
- ¹⁶ Mészáros, P., *Astr. Astrophys.*, **44**, 59–68 (1975).
- ¹⁷ Herterich, K., *Nature*, **250**, 311–313 (1974).
- ¹⁸ Colgate, S. A., *Astrophys. J.*, **150**, 163–192 (1967).
- ¹⁹ Arons, J., Kulsrud, R. M., and Ostriker, J. P., *Astrophys. J.*, **198**, 687–707 (1975), and references cited therein.
- ²⁰ Eachus, L. J., and Liller, W., *Astrophys. J. Lett.*, **200**, L61–L62 (1975).
- ²¹ Mészáros, P., *Nature*, **258**, 583–584 (1975).
- ²² Ilarionov, A. F., and Sunyaev, R. A., *Astr. Astrophys.*, **39**, 185–195 (1975).
- ²³ Shapiro, S. L., and Lightman, A. P., *Astrophys. J.*, **204**, 555–560 (1976).
- ²⁴ Charles, P. A., Longair, M. S., and Sanford, P. W., *Mon. Not. R. astr. Soc.*, **170**, 17P–22P (1975).
- ²⁵ Sanford, P. W., and Ives, J. C., *Proc. R. Soc. (in the press)*.

Source of terrestrial non-thermal radiation

Dyfrig Jones

Space Science Department, European Space Research and Technology Centre, Noordwijk, The Netherlands

Observations from satellites have revealed the existence of continuum radiation from just beyond the plasmopause. I propose here that this is generated by Cerenkov radiation coupling to the O mode of radiation, which can propagate in low density plasmas. The observations are interpreted on the basis of this theory.

RECENT observations by spacecraft of radio noise at distances beyond about six Earth radii (R_E) from the Earth indicate the presence of non-thermal continuum radiation (ref. 1, and D. A. Gurnett, unpublished) in the frequency range 500 Hz–100 kHz. Direction finding measurements on the Imp 6 spacecraft show that the radiation is generated in the outer radiation zone immediately beyond the plasmopause boundary¹. The source region extends from 0400 to 1400 (LT) and covers a broad latitude range, which includes the geomagnetic equator (D. A. Gurnett, unpublished); a separate source exists on the morning side of the magnetosheath.

The radiation has a sharp lower frequency cut-off near to the local electron plasma frequency, f_p , which is proportional to the square root of the electron density in the vicinity of the spacecraft; f_p ranges from 500 Hz in the distant magnetotail to > 100 kHz near the plasmopause. No distinct high frequency cut-off is observable and the intensity decreases as the frequency increases; > 100 kHz the radiation disappears below the cosmic noise level. The continuum seems to consist of two components, one which fills the whole magnetosphere and magnetotail and which is permanently trapped between the plasmasphere and magnetosheath, and another component which can penetrate the magnetosheath to propagate freely into the solar wind^{2,3}.

A spectrogram showing the low frequency cut-off of the radiation when Imp 6 was situated at $\sim 8R_E$ and 0545 (LT) is seen in Fig. 1, where the picture intensity is proportional to the amplitude of the noise. The signals from the electric and magnetic antennae are shown alternately; the magnetic signal appears weaker because the experiment sensitivity to this component is less.

The radiation is seen to have two distinct cut-offs, one at frequency $f_1 \approx 5.3$ kHz and the other at $f_2 \approx 6.35$ kHz. In the range between f_1 and f_2 the noise varies periodically as the spacecraft rotates, the nulls in the observed electric component occurring when the electric field vector of the radiation is perpendicular to the dipole antenna used for reception. It is stated that this electric field is polarised predominantly parallel to the geomagnetic field¹. At frequencies $> f_2$, bands of noise are apparent at ~ 7.7 kHz, 8.6 kHz and 9.4 kHz. These bands are not always observed but seem to appear irregularly when the spacecraft is close to the plasmopause. Since they are observed on both electric and magnetic channels they are apparently electromagnetic in character.

Mechanisms which have been proposed as a source of the continuum radiation are gyrosynchrotron radiation³ and electrostatic instabilities at half harmonics of the electron gyrofrequency¹, that is at $(n + \frac{1}{2})f_g$ where n is an integer and f_g is the gyrofrequency (which is proportional to the magnetic field strength). There is recent evidence (D. A. Gurnett, unpublished) that the radiation is dependent on the flux of particles of energy < 30 keV and this tends to eliminate the gyrosynchrotron mechanism which requires particles of much higher energy, ~ 1 MeV. The second mechanism proposed—electrostatic instabilities—requires that the energy be transferred

from electrostatic waves to electromagnetic waves and it is suggested that this can occur if the medium is inhomogeneous.

I here propose that the observed radiation is produced originally by the Cerenkov mechanism as electromagnetic waves in a mode that cannot normally propagate over large distances in the magnetosphere. Close to the source region, however, these waves may become coupled to electromagnetic waves in another mode which can propagate over large distances.

Theory

Cerenkov radiation is produced when particles are moving through a medium with velocities greater than the phase velocity of the radiation in the same direction. Since particle velocities cannot exceed the velocity, c , of electromagnetic radiation in free space, the Cerenkov criterion requires that the wave refractive index be greater than unity; the lower the speed of the particle, the larger the refractive index required.

Figure 2 illustrates the variation of the square of the refractive index (n^2) with X , when $Y < 1$ and collisions and temperature effects are neglected⁴; the parameters X and Y are as normally defined in magnetoionic theory⁵

$$X = f_p^2/f^2; \quad Y = f_g/f$$

where f is the wave frequency and f_p, f_g are the plasma frequency and gyrofrequency respectively. The condition $Y < 1$ requires that $f > f_g$; this condition and the assumption that collisions can be neglected are satisfied for wave frequencies > 5 kHz at radial distances of $\geq 6R_E$.

It is apparent from Fig. 2 that three modes of propagation are possible for the range of X shown. The mode whose refractive index goes to zero at (the mode with a cut-off at) $X = 1 - Y$ is the first extraordinary mode—the X mode. The mode with a cut-off at $X = 1$ is the ordinary (O) mode, and the mode having a cut-off at $X = 1 + Y$ is the second extraordinary mode, usually termed the Z mode. The shaded regions indicate where the curves of n^2 must lie as the angle θ between the wave normal direction and the magnetic field direction is varied; the full lines labelled a, b and c show the curves corresponding to some wave-normal angle θ intermediate to 0 and $\pi/2$.

It is important to note that as $\theta \rightarrow 0$, when the wave normal and magnetic field directions tend to be parallel (or antiparallel),

Fig. 1 v.l.f. Spectrogram obtained on Imp 6 (from ref. 1). Arrows indicate electric antenna parallel to geomagnetic field.

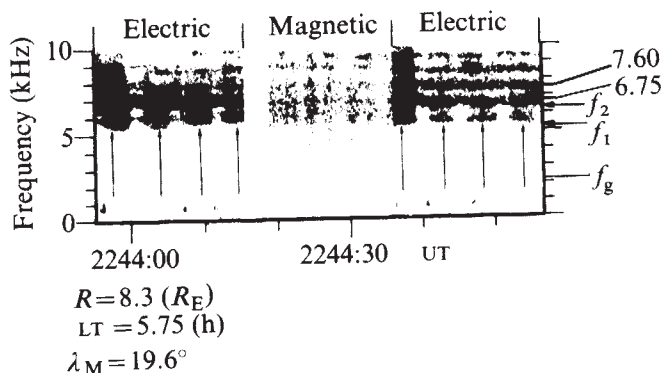
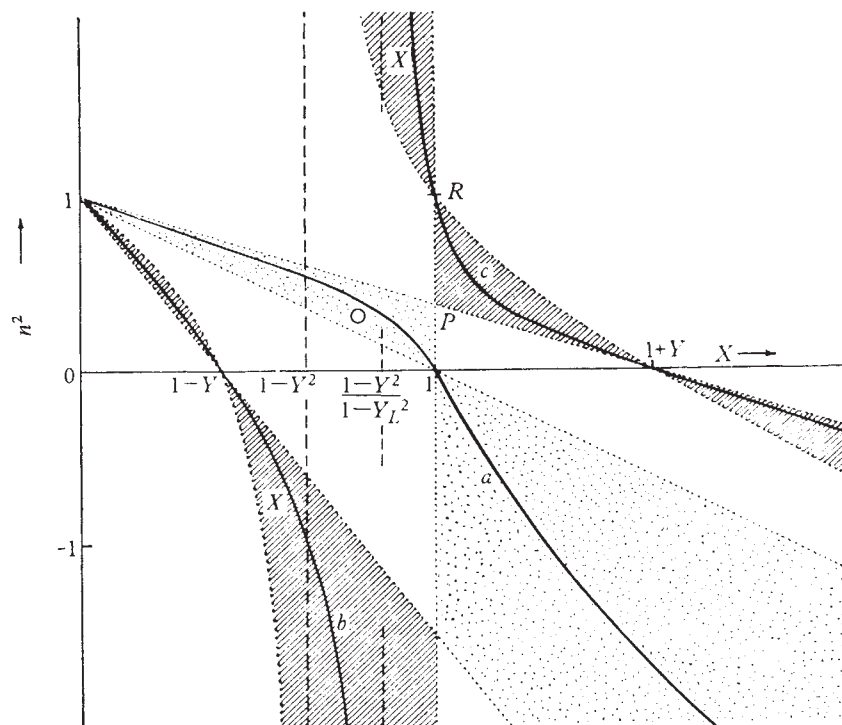


Fig. 2 Variation of (refractive index)² with X when $Y < 1$ (after Budden⁴). Electron collisions are neglected.



the curves a and c approach the point P at $X = 1$. The refractive index for the Z mode increases rapidly $< X = 1$ and $\rightarrow \infty$ at $X = (1 - Y^2)/(1 - Y_L^2)$ where $Y_L = Y \cos \theta$. As $\theta \rightarrow 0$, $Y_L \rightarrow Y$ and the infinity in $n \rightarrow X = 1$; as $\theta \rightarrow \pi/2$, $Y_L \rightarrow 0$ and the infinity $\rightarrow X = 1 - Y^2$. This infinity is termed the upper hybrid resonance (UHR).

The Cerenkov criterion can be satisfied only in the region $1 - Y^2 < X < 1$ since it is only there that $n > 1$. Observations on rockets and satellites in the ionosphere and magnetosphere⁶⁻¹⁶ show that under certain conditions, enhanced radiation, which has been called UHR noise, does exist in this frequency range extending from the plasma frequency $f = f_p$ ($X = 1$), to the UHR frequency $f = (f_p^2 + f_g^2)^{1/2}$. Calculations show that Cerenkov radiation from typical magnetospheric energy spectra is adequate to provide the power observed in UHR noise¹⁷. With reference to Fig. 2, it is clear that UHR noise cannot escape directly to regions of lower plasma density (smaller X) since the radiation is absorbed at the UHR. If, however, the Z waves propagate first into regions of increasing plasma density and reach the level $X = 1$ where the wave frequency becomes equal to the plasma frequency, then it is possible in certain conditions for the energy to be transferred or 'coupled' from the Z mode to the O mode. The O -mode energy is then free to propagate to lower values of X as is apparent from curve a in Fig. 2.

The theory of coupling⁴ yields a coupling parameter ψ given by the expression

$$\psi = -\frac{iY_T^2 Y_L}{4} \frac{X^1 + iZ^1}{(1 - X - iZ)^2 Y_L^2 + (Y_T^4/4)} \quad (1)$$

where $Y_T = Y \sin \theta$ and Z is the ratio of collision frequency to wave frequency and must not be confused with the Z mode. The parameters X^1 and Z^1 are the spatial gradients of X and Z respectively. It is seen from equation (1) that ψ can become large if simultaneously $X \rightarrow 1$ and $Z \rightarrow (1/2) |Y_T^2/Y_L|$. These are the conditions for 'critical coupling'. In the magnetosphere beyond a few thousand kilometres altitude, the value of Z computed on the basis of Coulomb collisions is small and thus critical coupling is achieved if $Y_L \gg Y_T$ at $X = 1$. Therefore, one can expect energy transfer from the X mode to the O mode near $X = 1$ if the wave normal direction is close to the magnetic

field line direction; in this condition curves a and c in Fig. 2 approach one another at point P .

Further theory

To understand in more detail the emission process and the consequent wave propagation and coupling, it is necessary to consider the refractive index surfaces (n surfaces). An n surface is a surface on which lies the tip of a vector whose direction is that of the wave normal and whose magnitude is proportional to the phase refractive index for that direction. The n surfaces for the Z mode are illustrated in Fig. 3a. For a given value of Y , a family of n surfaces may be drawn corresponding to different values of X . The surfaces are in fact surfaces of revolution about the geomagnetic field $\mathbf{B}_0$.

To determine graphically the wave normal angle at which Cerenkov radiation is produced by a particle with velocity $c\beta$ along the magnetic field, $\beta < 1$, a line may be drawn normal to $\mathbf{B}_0$ at a distance $1/\beta$ from the origin. Where this line intersects an n surface—for example at M in Fig. 3a—the Cerenkov condition is satisfied; the radiation produced at the level corresponding to $X = 0.9$ will then have its wave normal along $\mathbf{OM}$. Note that for $X < 1$ the shape of the n surfaces is such that particles of all energies (all values of β) will produce waves all of which have nearly the same wave normal direction. In practice, temperature effects limit n to finite values, so that the lower energy particles do not contribute to the power radiated. After the waves are produced, propagation into regions of lower or higher X may be followed by using Poeverlein's construction. This is based on Snell's law, which requires that the tangential components n be equal above and below an isoionic plane (plane of constant X) through which the wave travels⁷. In Fig. 3a consider the isoionic plane AA' which has been selected so that its normal passes through points M and P ; the significance of this will appear later. The original wave normal direction is along $\mathbf{OM}$ but the group velocity of the wave is in the direction perpendicular to the n surface at M (see Budden⁴) as indicated by the small arrow; this is the ray direction. If the energy propagates into regions of increasing electron density (increasing X) the locus of the tip of the wave normal moves from n surface to n surface along the line MN . The ray path resulting from the condition that the ray direction should

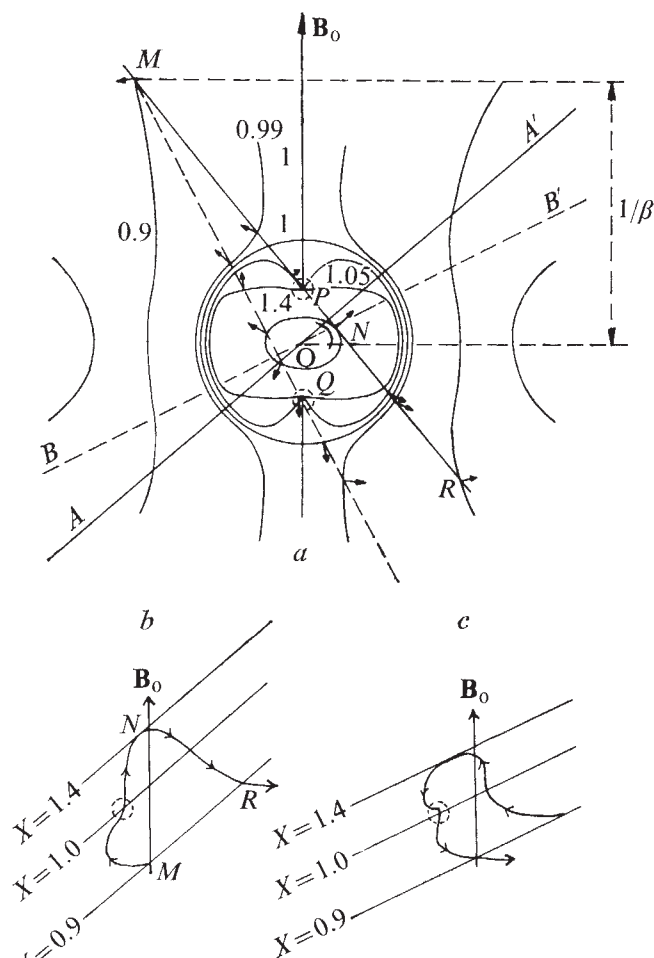


Fig. 3 a, Refractive index surfaces for the X-wave when $Y < 1$. Figures on the curves are values of X . b, Ray path with isoionic contours AA' . c, Ray path with isoionic contours BB' . The dotted circles indicate the regions of coupling.

always be normal to the n surfaces at the tip of the wave normal is shown in Fig. 3b. At point N the wave is reflected, its wave normal following the line NR in Fig. 3a and its ray following NR in Fig. 3b.

It is evident from Fig. 3a that when the waves pass through the region corresponding to $X = 1$, the wave normal becomes aligned with B_0 and the two conditions for critical coupling are satisfied. Energy can be transferred at this point from the Z wave to the O wave whose n surface at this level also passes through point P as can be deduced from Fig. 2 (see also Budden⁴). The O wave so produced cannot propagate into regions of higher X (see Fig. 1) but is immediately reflected to propagate unhindered to regions of lower plasma density. A similar process occurs if the isoionic planes are aligned parallel to BB' , except that in this case the coupling occurs near $X = 1$ after the Z wave has been reflected; the corresponding ray path is shown in Fig. 3c.

The significance of the alignment of AA' and BB' is now clear—these are the orientations of the isoionic planes which render the coupling most efficient. It is seen from equation 1 that the coupling parameter ψ is also dependent on X^1 and therefore if X^1 is large one may expect some coupling to occur, even when the isoionic planes are not strictly aligned to AA' or BB' . Calculating the actual amount of coupling for different wave normal angles is the subject of further study.

Interpretation of observations

On the basis of the mechanism proposed, the frequency cut-offs appearing in the Imp 6 spectrogram (Fig. 1) can be identified.

The lower cut-off at f_1 corresponds to the Z -mode cut-off ($X = 1 + Y$) at the spacecraft; frequencies $< f_1$ will not reach the spacecraft, since they encounter cut-offs in the lower density regions above the spacecraft altitude. In fact, if f_1 is to correspond exactly to the condition $X = 1 + Y$ at the spacecraft, some waves of this frequency must be incident normally on the isoionic plane, since waves incident obliquely will be reflected before reaching $X = 1 + Y$. The Cerenkov radiation is expected to be produced with a fairly large range of wave normal angles since usually a wide range of particle energies is involved; the effect on the wave normal angle of varying β may be deduced from Fig. 3a. It is expected, therefore, that there will be a sufficiently broad distribution of wave normal angles to ensure that some of the waves come very close to the level $X = 1 + Y$ before being reflected.

The other frequency cut-off at f_2 may be identified with the level $X = 1$ where the O wave is produced. At this level the O waves will have their normals mainly aligned to the geomagnetic field, this being a condition for their production. Thus one does not have a broad distribution of wave normal angles as for the Z mode and, theoretically, reflection of the O wave will occur before the level $X = 1$ is reached for the case shown in Fig. 3b. It is probable, however, that the region of coupling is so close to $X = 1$ that the cut-off at f_2 can be attributed to the O -wave cut-off at $X = 1$. This is also true if the coupling occurs after X -wave reflection (Fig. 3c).

From the identification of the cut-off frequency $f_2 = f_p = 6.35$ kHz, the plasma density at the spacecraft is calculated to be 0.5 electrons cm^{-3} , and if one uses the relation $X = f_p^2/f_1^2 = 1 + Y = 1 + f_g/f_1$, the electron gyrofrequency f_g at the spacecraft is found to be 2.3 kHz. The value of f_g at the spacecraft calculated from a harmonic expansion of the geomagnetic field is 2.2 kHz. A previous identification¹ assumed f_1 to be the plasma frequency and f_2 to correspond to the X -wave cut-off at $X = 1 - Y$ (see Fig. 2). This resulted in an estimate of electron density of 0.35 electrons cm^{-3} and a gyrofrequency of 1.9 kHz. The maximum possible error from such a misidentification would amount to a factor of 2 in plasma density.

In the frequency range between f_1 and f_2 , the electric field of the radiation should be predominantly perpendicular to the static magnetic field with the longitudinal component increasing rather rapidly near the plasma frequency^{18,19}. With a single dipole antenna it is difficult to determine unambiguously the orientation of the electric field vector when the wave is linearly polarised, and a more detailed investigation of the spin-modulated noise between f_1 and f_2 using three mutually orthogonal antennae is required.

From the new identification of cut-offs presented here, the UHR frequency at the spacecraft is calculated to be $f_{\text{UHR}} = 6.75$ kHz and the cut-off of the X wave should be at $f_X = 7.6$ kHz. These frequencies are indicated in Fig. 1, and they are seen to correspond to abrupt changes in noise intensity. The larger intensity < 6.75 kHz is understandable since it is in this range $f_2 < f < f_{\text{UHR}}$ that the noise is believed to be produced by the Cerenkov mechanism. The fact that the intensity change occurs very close to the UHR, indicates that these waves have large wave normal angles, otherwise the intensity would have changed at a lower frequency corresponding to $X = (1 - Y^2)/(1 - Y_L^2)$ as may be deduced from Fig. 2.

It is tempting to identify the increase in intensity > 7.6 kHz as being due to the presence of the X wave. There is no known mechanism for producing this mode locally but it is possible that it could have been formed by depolarisation of the O wave by reflections at the magnetosheath and at magnetospheric irregularities in plasma density. The resulting X waves could then propagate inwards and be reflected at the level $X = 1 - Y$, thus producing an abrupt change in noise intensity at 7.6 kHz. The noise bands at higher frequencies would, however, remain unexplained.

An explanation of the noise bands usually observed close to the plasmopause as being harmonics of the local gyrofrequency has been ruled out, since the difference in frequency between

bands in Fig. 1 is ~ 1 kHz whereas the gyrofrequency is > 2 kHz. It also seems that the noise bands are not accurately harmonically related. One possible explanation for these bands is that they are produced as a consequence of ionisation ducts below the spacecraft altitude. As is seen from equation 1, plasma density gradients within ducts would render coupling more efficient, X^1 being larger, and one could expect bands of more intense radiation at frequencies corresponding to the duct plasma densities. On this assumption and by using an average plasma density profile¹ for the magnetosphere, one can estimate the distance between ducts below the spacecraft to be $\sim 0.3R_E$. It is significant that ducts are known to exist predominantly within the plasmasphere and the noise bands are usually seen close to the plasmapause.

Conclusions

It is concluded that mode coupling of Cerenkov radiation may be the source of terrestrial non-thermal radiation. The mechanism requires that the plasma be traversed by energetic particles and this is known to be the case in the magnetosphere. The angle between the geomagnetic field and the isoionic planes is an important parameter and this may explain the local-time asymmetry in the source location, since it is well known that the plasmasphere has a distinct outward bulge at local evening. In addition, the dawn–dusk magnetospheric electric field increases the local–morning electron energies (D. A. Gurnett, unpublished) by ~ 10 keV and this has obvious implications on the Cerenkov mechanism proposed.

The radiation which escapes to regions of lower plasma density

will be in the ordinary mode which is left-hand polarised; polarisation measurements of the noise are not yet available. The radiation whose frequency is greater than the magnetosheath plasma frequency will escape into the solar wind; the radiation whose frequency is less than the sheath plasma frequency will be reflected and will be trapped between the higher plasma densities of the sheath and plasmasphere, ultimately escaping down into the magnetotail.

Conditions suitable for the mechanism proposed also exist in the morning magnetosheath, which is a favoured region for streaming electrons from the solar wind^{20,21}.

Received June 9, 1975; accepted February 2, 1976.

- ¹ Gurnett, D. A., and Shaw, R. R., *J. geophys. Res.*, **78**, 8136–8149 (1973).
- ² Brown, L. W., *Astrophys. J.*, **180**, 359 (1973).
- ³ Frankel, M. S., *Radio Sci.*, **8**, 991 (1973).
- ⁴ Budden, K. G., *Radio Waves in the Ionosphere* (Cambridge University Press, 1966).
- ⁵ Ratcliffe, J. A., *The Magnetoionic Theory and Its Applications to the Ionosphere* (Cambridge University Press, 1962).
- ⁶ Walsh, D., Haddock, F. T., and Schulte, H. F., *Space Res.*, **4**, 935 (1964).
- ⁷ Harvey, C. C., *Annals Astrophys.*, **28**, 248 (1965).
- ⁸ Bauer, S. J., and Stone, R. G., *Nature*, **218**, 1145 (1968).
- ⁹ Harvey, C. C., *Nature*, **217**, 50 (1968).
- ¹⁰ Gregory, P. C., *Nature*, **221**, 350 (1969).
- ¹¹ Barrington, R. E., and Hartz, T. R., *Plasma Waves in Space and Lab.*, **1**, 55 (Edinburgh University Press, 1969).
- ¹² Hartz, T. R., *Proc. IEEE*, **1042** (1969).
- ¹³ Hartz, T. R., *The Polar Ionosphere and Magnetospheric Processes* (Gordon and Breach, New York, 1970).
- ¹⁴ Muldrew, D. B., *Space Res.*, **10**, 786 (1970).
- ¹⁵ Barrington, R. E., Hartz, T. R., and Harvey, R. W., *J. geophys. Res.*, **76**, 5278 (1971).
- ¹⁶ Mosier, S. R., Kaiser, M. L., and Brown, L. W., *J. geophys. Res.*, **78**, 1673 (1973).
- ¹⁷ Taylor, W. W. L., and Shawhan, S. D., *J. geophys. Res.*, **49**, 105 (1974).
- ¹⁸ Harvey, C. C., thesis, Univ. Cambridge, (1965).
- ¹⁹ Budden, K. G., and Hugill, J., *Proc. R. Soc. A*, **277**, 365 (1964).
- ²⁰ Scarf, F. L., Fredricks, R. W., Frank, L. A., and Neugebauer, M., *J. geophys. Res.*, **76**, 5162 (1971).
- ²¹ Fredricks, R. W., Scarf, F. L., and Frank, L. A., *J. geophys. Res.*, **76**, 6691 (1971).

letters to nature

Type I supernovae and galactic production of iron

CURRENT models of galactic chemical evolution (see ref. 1) assume that heavy elements are produced by massive stars, that is, in type II supernova events. We show here that Fe may have a different source from other heavy elements—in type I supernovae which result from the deaths of moderately long lived stars.

Type I and type II supernovae both have similar radii and temperatures near maximum light², yet the spectra of type I supernovae show considerably more structure. Many of the lines have been attributed to Fe³. At a time long after maximum light, Kirshner and Oke⁴ were able to study the emission from SN 1972e in NGC 5253 in a quantitative way. If their identification of Fe II lines is correct, $\sim 0.03M_{\odot}$ of Fe is required. This is a lower limit to the Fe abundance and provides strong observational evidence for Fe production in type I supernovae.

Observations of type II supernovae at around maximum light do not show any overabundances of heavy elements. This is easily understood since the envelope of the presupernova star is being observed. The only observable material which has been recently processed by a massive star may be the fast moving knots in Cas A (refs 5, 6). Only lines of Ar, O, and S have been detected; no Fe lines have been observed although Fe II lines might be expected. No attempt has yet been made to put a limit on the Fe abundance. Thus, while there is no observational evidence for Fe production in type II supernovae, there is no evidence to the contrary. Stellar evolution calculations do not clarify the problem since it is unknown how much Fe is contained in the possible compact remnant of a supernova.

The last piece of evidence on Fe production comes from the theory of supernova light curves. First consider the supernova energy to be deposited instantaneously. If the observed supernova luminosity near maximum light is to be reproduced, the radius of the presupernova star must be $\geq 10^{14}$ cm (refs 7, 8). This condition is satisfied for massive stars, but may not be the case for the small mass progenitors of type I supernovae. If the initial radius is small, some supernova energy must be stored over tens of days. Colgate and McKee⁸ have suggested that the energy is stored in the radioactive products of explosive nucleosynthesis. The model which reproduced the observed magnitude of type I supernovae produced $0.25M_{\odot}$ of Fe. If any more Fe is produced explosively, the models predict excessive luminosity. It should be noted that the observed luminosity being reproduced in these models is based on a Hubble constant, H , of 100. A lower value of H would require greater Fe production. Thus, from observations and models of type I supernovae, their Fe yield is in the range of 0.03 – $0.25M_{\odot}$.

This yield can be compared with that required to give the present solar abundance of Fe. Using solar neighbourhood values for the birthrates of moderate-mass stars and galactic mass density, Ostriker *et al.*⁹ derive limits on the Fe production per massive star. It is certainly $< 0.23M_{\odot}$ of Fe per star and is very likely to be $< 0.023M_{\odot}$ of Fe. The birthrate would be equivalent to the type II supernovae rate. Tammann¹⁰ finds that types I and II supernovae have similar total rates and radial distributions in galaxies like our own. Thus the yield found by Ostriker *et al.* also applies to type I yields.

These considerations indicate that even the observed amount of Fe in type I supernovae may be sufficient to give the Fe

abundance in the solar neighbourhood. If the low Fe production per supernova is correct, the radioactivity model for type I supernovae is incorrect, and models with either extended initial envelopes or another form of late energy input must be considered.

Fe production in type I supernovae has implications for the chemical history of the Galaxy. It is generally accepted that type I supernovae result from less massive stars than those giving type II supernovae (for a recent review, see Tinsley¹¹). The lifetimes of the pre-type I supernovae stars may be an order of magnitude longer than that of the pre-type II supernovae stars. Assume that the stellar mass function is independent of time. During the initial burst of star formation, there is a lag between the production of Fe in type I supernovae and the production of the other heavy elements in type II supernovae. During this period, stars can be formed which are underabundant in Fe relative to the other heavy elements. There is some evidence that old stars are overabundant in O relative to Fe (ref. 12), but the observed abundances may not reflect the initial composition of these stars. Peimbert¹³ notes that the O overabundance may well be *primaevae*. He also reviews evidence that at least two parameters are required to describe the horizontal branch characteristics of globular clusters and suggests that the two parameters are the Fe abundance and the CNO abundance.

If considerable heavy element processing occurs in the disks of spiral galaxies, the type I supernova origin of Fe is of some importance. Observations of supernovae show that type II supernovae are concentrated in spiral arms while type Is are scattered throughout the disk¹⁴. A given volume of gas thus receives processed material from type II supernovae in bursts and material from type Is more uniformly. The result is that the ratio of Fe to other heavy elements may be expected to show fluctuations.

Heavy elements are overabundant in cosmic rays in a way that is consistent with the supernova origin of cosmic rays. Ramaty *et al.*¹⁵ found that the energy spectrum of Fe cosmic rays is flatter than that of CNO nuclei. They interpret the result as indicating that Fe nuclei are accelerated near to a pulsar while the other heavy elements are accelerated in the surrounding nebula. The present work suggests a different interpretation: Fe nuclei are accelerated in type I supernovae and CNO nuclei are accelerated in type II supernovae. If a Fermi-type acceleration mechanism is operating, the flatter Fe energy spectrum may arise from the higher material velocities in type I supernovae.

In the future, observational checks on the source of galactic Fe will be possible. Clayton¹⁶ has reviewed arguments that Fe is produced explosively and claims that γ -ray line observations of supernovae will enable abundances of nucleosynthesis products to be determined. Also, young supernova remnants such as the remnant of Tycho's supernova of 1572 should be examined for the presence of Fe emission. The supernova ejecta may have been heated to high temperature in a reverse shock wave or a conduction front, so that X-ray observations with high spectral resolution may be the most profitable. Third, observations of extragalactic supernovae, particularly at late times, will be useful. These observations should yield information on whether type I supernovae are important contributors of galactic iron.

The Kitt Peak National Observatory is operated by the Association of Universities for Research in Astronomy Inc., under contract with the NSF.

ROGER A. CHEVALIER

Kitt Peak National Observatory,
Tucson, Arizona 85726

Received December 24, 1975; accepted March 16, 1976.

- ⁶ Arnett, W. D., *Astrophys. J.*, **195**, 727 (1975).
- ⁷ Colgate, S. A., and White, R. H., *Astrophys. J.*, **143**, 626 (1966).
- ⁸ Colgate, S. A., and McKee, C., *Astrophys. J.*, **157**, 623 (1969).
- ⁹ Ostriker, J. P., Richstone, D. O., and Thuan, T. X., *Astrophys. J. Lett.*, **188**, L87 (1974).
- ¹⁰ Tammann, G. A., in *Supernovae and Supernova Remnants* (edit. by Cosmovici, C. B.), 155 (Reidel, Dordrecht-Holland, 1974).
- ¹¹ Tinsley, B. M., *Publ. astr. Soc. Pacif.*, **87**, 837.
- ¹² Conti, P. S., Greenstein, J. L., Spinrad, H., Wallerstein, G., and Vardya, M. S., *Astrophys. J.*, **148**, 105 (1967).
- ¹³ Peimbert, M., in *The Formation and Dynamics of Galaxies* (edit. by Shakeshaft, J. R.), 141 (IAU Symp., **58**, 1974).
- ¹⁴ van den Bergh, S., and Maza, J., *Astrophys. J.* (in the press).
- ¹⁵ Ramaty, R., Balasubrahmanyan, V. K., and Ormes, J. F., *Science*, **180**, 731 (1973).
- ¹⁶ Clayton, D. D., in *Explosive Nucleosynthesis* (edit. by Schramm, D. N., and Arnett, W. D.), 264 (University of Texas Press, Austin, 1973).

X-ray emission from γ Cas

THE Copernicus orbiting observatory contains two astronomical experiments, the Princeton ultraviolet spectrometer and the MSSL X-ray package. They are coaligned, and during ultraviolet observations the 3–8 keV collimated proportional counter routinely collects background data¹. When the observatory was pointed at the ultraviolet target γ Cas on December 21, 1975 and again on January 20, 1976, a clear excess of counts was seen in the 3–8 keV channel compared to preceding and following targets. This count rate corresponds to a flux at the Earth of $\sim 1.2 \times 10^{-10}$ erg cm⁻² s⁻¹, and lies 5σ above the mean background level determined from several weeks of observations (Table 1).

We have also examined earlier Copernicus X-ray data taken on γ Cas. The star was observed on November 4, 1973 and on November 13, 1974. No flux was detected during the 1973 observation, but the sensitivity attainable was low because of the sparsity of background data near that time. A signal was detected in 1974, 4σ above the mean background.

The FWHM field of view of the Copernicus X-ray detector is $\sim 3''$, and we cannot localise the emission to greater accuracy than this. Jernigan² has, however, reported that the SAS-3 satellite, in a 2-d observation beginning December 30, 1975, observed X-ray emission of comparable intensity to that reported here from a region of sky within $1'$ of γ Cas. This suggests a high probability that the X-ray emission originates from the vicinity of this star, although we note that γ Cas does have a companion ADS782B ($m_v \approx 10$), the two stars being separated by $\sim 2''$ (ref. 4). The X-ray source was not detected before November 1974; in particular, it was not recorded in the Uhuru sky survey^{3,4}, which suggests that it is variable. There is no evidence within the Copernicus data for flux variations $> 1\sigma$, but the strength reported by the SAS-3 group² (3.5×10^{-10} erg cm⁻² s⁻¹ from 2–11 keV) is a factor of ~ 1.5 above the highest value seen by Copernicus, allowing for the difference between the energy range of the two detectors.

γ Cas is an extreme Be star whose magnitude is variable between 1.6 and 3.0 (ref. 4). Its absolute visual magnitude is probably ~ -4 (ref. 4), so that the intrinsic luminosity of the X-ray source is $\sim 10^{32}$ erg s⁻¹. Maraschi, Treves and van den Heuvel⁵ have noticed the several coincidences

Table 1 X-ray data from γ Cas

Observations		Duration (h)	c.p.m. (3–8 keV)	erg cm ⁻² s ⁻¹ ($\times 10^{-10}$)
Mid-time				
1973 November	4.4	16	< 3.0	< 1.2
1974 November	13.7	37	2.0 ± 0.5	0.8
1975 December	21.6	12	3.3 ± 0.6	1.3
1976 January	20.4	22	2.8 ± 0.5	1.1

The uncertainties in counting rate are dominated by systematic background effects which are measured empirically. The energy flux is derived assuming a Crab-like spectrum, but the conversion factor from c.p.m. to erg cm⁻² s⁻¹ is relatively insensitive to spectral slope and form.

¹ Talbot, R. J., and Arnett, W. D., *Astrophys. J.*, **170**, 409 (1971).

² Kirshner, R. P., Oke, J. B., Penston, M., and Searle, L., *Astrophys. J.*, **185**, 303 (1973).

³ Branch, D., and Patchett, B., *Mon. Not. R. astr. Soc.*, **161**, 71 (1973).

⁴ Kirshner, R. P., and Oke, J. B., *Astrophys. J.*, **200**, 574 (1975).

⁵ Peimbert, M., *Astrophys. J.*, **170**, 261 (1971).

between Be stars and X-ray sources. In particular, γ Cas is very similar in its optical characteristics to X Per (ref. 4), which has been associated with the regularly varying X-ray source 3U0352+30 (refs 6–8). The ratio of X-ray to optical luminosity of the two stars now differs by just over an order of magnitude⁸. No evidence is found for regular modulation in the present data between 2 and 40 min. The source is, however, comparatively weak, and any modulation of peak to mean amplitude <60% of the mean flux would have gone undetected. More sensitive observations will be necessary to explore further the possible parallels between γ Cas and X Per. Meanwhile, γ Cas provides yet another example of the growing number of intrinsically relatively weak X-ray sources being discovered which are associated with nearby stellar objects (see ref. 9).

We are grateful to P. A. Charles for discussions. K.O.M. and N.E.W. acknowledge support from the SRC.

K. O. MASON

Mullard Space Science Laboratory,
University College London, and
Mullard Radio Astronomy Observatory,
Cavendish Laboratory, Cambridge, UK

N. E. WHITE
P. W. SANFORD

Mullard Space Science Laboratory,
University College London, UK

Received February 12; accepted March 18, 1976.

- ¹ Sanford, P. W., *Proc. R. Soc.*, A340, 411 (1974).
- ² Jernigan, J. G., *IAU Circ.* No. 2900 (1976).
- ³ Giacconi, et al., *Astrophys. J. Suppl.*, 27, 37 (1974).
- ⁴ Moffat, A. F. J., Haupt, W., and Schmidt-Kaler, Th., *Astr. Astrophys.*, 23, 433 (1973).
- ⁵ Maraschi, L., Treves, A., and van den Heuvel, E. P. J., *Nature*, 259, 292 (1976).
- ⁶ Hawkins, F. J., Mason, K. O., and Sanford, P. W., *Astrophys. Lett.*, 16, 19 (1975).
- ⁷ White, N. E., Mason, K. O., and Sanford, P. W., *IAU Circ.* No. 2854 (1975).
- ⁸ White, N. E., Mason, K. O., Sanford, P. W., and Murdin, P., *Mon. Not. R. astr. Soc.* (in the press).
- ⁹ Catura, R. C., Acton, L. W., and Johnson, H. M., *Astrophys. J. Lett.*, 196, L47 (1975).

Radiometric diameter and albedo of the remarkable asteroid 1976AA

THE minor planet 1976AA (Fast-Moving Object Helin), discovered on January 7, 1976, is the first asteroid found with an orbital period of < 1 yr (orbital elements $a=0.97$, $e=0.18$; $i=19^\circ$; ref. 1). Although it seems to be a likely speculation that 1976AA is a former Apollo object that originally had its aphelion in the main asteroid belt, it is possible that this object is a member of a group of minor planets that is distinct in physical as well as orbital properties. Observations that reveal the physical nature (colour, albedo, texture, mineralogy) of its surface are needed to compare it with better studied objects in more orthodox orbits. This letter reports a determination by infrared radiometry of the diameter and albedo of 1976AA.

The radiometric technique of measuring diameters and albedos was first applied to asteroids in 1970 (ref. 2) and has recently been discussed in some detail^{3,4}. Physically, it is based on the dependence on albedo of the balance between reflected and thermally emitted radiation for an airless body in thermal equilibrium with the sunlight striking it. The application reported here follows the formulation and calibration by Jones and Morrison⁴, which has previously been used to determine the diameters of > 60 main-belt asteroids^{5,6} and the Earth-approaching object 433 Eros⁷.

The observations were made on January 26, 1976 in a broad spectral band centred at $\sim 10 \mu\text{m}$ with the infrared photometer⁸ of the 1.5-m telescope of the Lunar and Planetary Laboratory Catalina Observatory. The primary stellar standards were α Tau and β Gem, for which we adopted magnitudes of -3.1 and -1.3 , consistent with the

standard calibration assumed for the radiometric diameter technique⁹. Two groups of integrations, centred at UT dates 26.22 and 26.30, yielded a $10\text{-}\mu\text{m}$ magnitude of $+5.0 \pm 0.3$. At the time of observation, 1976AA was 0.150 AU from Earth and 1.075 AU from the Sun, and its phase angle was 50° . Although it is not clear exactly how the radiometric brightness should be reduced to full phase, both model calculations and observations of 433 Eros⁷ suggest a correction of between -0.01 and $-0.02 \text{ mag deg}^{-1}$, yielding an equivalent $10\text{-}\mu\text{m}$ magnitude at zero phase of $+4.3 \pm 0.5$. The value of $V(1,0)$, the zero-phase, unit-distance visual magnitude, is $+17.6$ from other observations made here (J.C.G., unpublished).

A straightforward solution for diameter and visual geometric albedo based on these observations yields $D=900 \pm 200 \text{ m}$ and $p_V=0.20 \pm 0.07$. This albedo is in satisfactory agreement with a preliminary polarimetrically derived value (J. C. Gradie, unpublished), and the diameter is the smallest ever measured for an asteroid.

On the basis of its albedo, 1976AA can probably be classed with the stony or siliceous (S-class) group of asteroids⁹. These asteroids seem to have related surface mineralogies and to be the predominant type near the inner edge of the belt, although the much darker carbonaceous objects constitute the great majority of all asteroids. The previously studied Mars- and Earth-orbit-crossing objects—433 Eros, 887 Alinda, 1566 Icarus, and 1685 Toro—also belong to the broad S compositional class. Thus, on the basis of this preliminary physical datum, 1976AA appears similar physically to other small asteroids that approach the Earth.

This research was supported in part by grants from NASA. We thank E. Helin for informing us of the existence of 1976AA, and J. G. Williams for providing the ephemeris without which these observations could not have been made.

DAVID MORRISON*
JONATHAN C. GRADIE
GEORGE H. RIEKE

Lunar and Planetary Laboratory,
The University of Arizona,
Tucson, Arizona 85721

Received February 10; accepted March 18, 1976.

*Permanent address: Institute for Astronomy, University of Hawaii, Honolulu.

- ¹ Helin, E., *IAU Circ.* No. 2899.
- ² Allen, D. A., *Nature*, 227, 158–159 (1970).
- ³ Morrison, D., *Icarus*, 19, 1–14 (1973).
- ⁴ Jones, T. J., and Morrison, D., *Astr. J.*, 79, 892–895 (1974).
- ⁵ Morrison, D., *Astrophys. J.*, 194, 203–212 (1974).
- ⁶ Morrison, D., and Chapman, C. R., *Astrophys. J.*, in the press.
- ⁷ Morrison, D., *Icarus* (in the press).
- ⁸ Low, F. J., and Rieke, G. H., in *Methods of Experimental Physics* 12, Part A (edit. by Carlson, N.), 415–462 (Academic, New York, 1974).
- ⁹ Chapman, C. R., Morrison, D., and Zellner, B., *Icarus*, 25, 104–130 (1975).

On simultaneous tilt and creep observations on the San Andreas Fault

THE installation of an array of tiltmeters along the San Andreas Fault¹ has provided an excellent opportunity to study the amplitude and spatial scale of the tilt fields associated with fault creep. We report here preliminary results from, and some implications of, a search for interrelated surface tilts and creep event observations at four pairs of tiltmeters and creepmeters along an active 20-km stretch of the San Andreas Fault. We have observed clear creep-related tilts above the instrument resolution (10^{-8} rad) only on a tiltmeter less than 0.5 km from the fault. The tilt events always preceded surface creep observations by 2–12 min, and were not purely transient in character.

Although episodic fault creep has been observed at many locations of the San Andreas, Hayward, Calaveras, and other active faults in the San Andreas Fault System^{2–8},

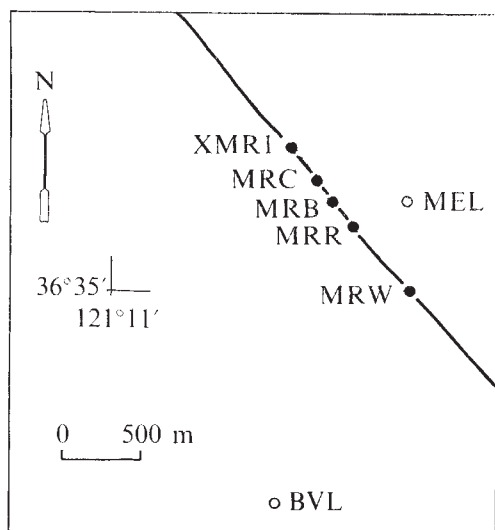


Fig. 1 Location map in the Bear Valley region of central California of tiltmeters (MEL and BVL) and creepmeters (MRW, MRR, MRB, MRC and XMR1) along the San Andreas Fault.

measurements of transient tilts and strains associated with creep events have not yet been made in sufficient detail to allow estimates of critical parameters such as depth, dimension, amount of slip, and the effects of material property variations. These parameters are critical to questions concerning, first, the degree to which the dynamic aseismic slip behaviour of the fault is reflected in fault creep measurements; second the relationship between observed creep and seismicity; and third, the source of the nonlinear form of the creep signals. Typical creep events have amplitudes of a few millimetres and durations of a few minutes to a few days^{8,9}. Fault displacements inferred from cumulative creep¹⁰ are in approximate agreement with those measured using geodetic techniques¹¹, and approximately simultaneous observations of creep events at closely spaced creepmeter sites have been interpreted as slip propagation along the fault at rates of up to 10 km d^{-1} (refs 4 and 6).

With one exception, the tiltmeters considered here form part of an array of shallow borehole instruments approximately 6 km apart, and 1-2 km from the fault (Fig. 1), intended primarily for the detection of earthquake related deformation¹. The exception is the tiltmeter at Melendy

Ranch (MEL) which was installed 370 m from the fault near several Melendy Ranch creepmeters. This instrument grouping, in addition to tiltmeter MEL on the north-east side of the fault, consists of creepmeters XMR1, MRC, MRB, MRR, and MRW along the fault from north-west to south-east, and tiltmeter BVL on the south-west side of the fault, as shown in Fig. 1. At these creepmeters clear examples of episodic creep on the San Andreas Fault are observed every few months^{8,9}.

Since January, 1974, three of the largest creep events, each recorded on at least two creepmeters, have occurred while both tiltmeters were operating. Figure 2 shows about 6 h of parallel record from the two tiltmeters and the particular creepmeters in operation at the time of these creep events. The uncertainty in timing for all sets of records is less than 2 min.

For the first creep sequence of July 11, 1974 (Fig. 2a) creep was recorded only on meters MRC and XMR1 to the north-west of MEL. The times of onset were 1745 and 1808 GMT, respectively. For more than an hour before 1743 a ramp-like change is evident in the north-south tilt component at MEL (Fig. 2a). Although interesting, this is the only example of a tilt ramp of this sort, and its meaning is not clear. From 1743 to 1800 the tilt at MEL increased by $0.5 \mu\text{rad}$ in a direction $\text{S}41^\circ\text{W}$. The amplitude decreased and the direction changed slowly clockwise during the next 30 min, finally resulting in a tilt of $0.4 \mu\text{rad}$ towards the west. No significant changes occurred in the tilt record at BVL (resolution $10^{-2} \mu\text{rad}$) during this time.

A second creep sequence, on October 3, 1974 (Fig. 2b), is more important since it was observed on the four creepmeters MRW, MRB, MRC, and XMR1 along the fault from the south and to the north-west of MEL. The first onset of creep occurred at 0812 at MRC. Almost simultaneous offsets occurred next at MRW and MRB at 0815 followed by XMR1 at 0837. The tilt changed at MEL between 0800 and 0815 by $0.52 \mu\text{rad}$ in a direction $\text{S}72^\circ\text{W}$, and during the next hour this amplitude decreased and the direction changed slightly clockwise until reaching a final tilt offset value of $0.36 \mu\text{rad}$ in a westerly direction. The tilt records at BVL were again unchanged.

The creep sequence on April 14, 1975 (Fig. 2c) was recorded on MRR at 0805, followed by MRC at 0821 and on XMR1 at 0826. Only the east-west tilt component was operating at MEL during this time. This record shows a tilt change of $0.25 \mu\text{rad}$ between 0800 and 0807 similar to that observed for the earlier creep sequences.

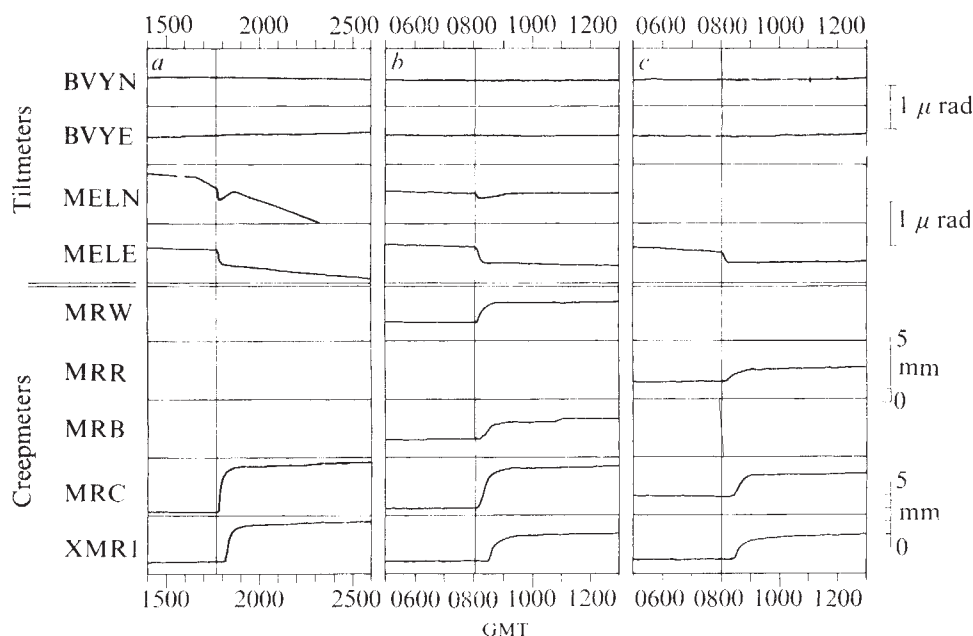


Fig. 2 Three examples of approximately 6 h of simultaneous digitised records from creepmeters XMR1, MRC, MRB, MRR and MRW and tiltmeters BVL and MEL during creep events on: a, July 11, 1974; b, October 3, 1974; c, April 14, 1975. BVYN, MELN, and BVLE, MELE are the north and east components, respectively. Absence of data indicates an instrument malfunction.

For the next pair of instruments to the north, creepmeter XFL1 and tiltmeter SAS with a 2-km separation, five creep sequences have occurred since June, 1973 when simultaneous recording began. Tiltmeter SAS is 1.2 km from the fault. At that distance, as for BVL, no clear relationship between the creep events and either short or intermediate term tilts has been found. This is true also for the northernmost instrument pair considered; creepmeter XPR1 and the tiltmeter LIB, separated by 3.4 km with LIB 1.5 km from the fault.

Creep-related tilts seem, therefore, to be rapidly attenuated as a function of distance from the fault. At Melendy Ranch, in particular, a displacement of a few millimetres on the fault that results in a tilt of $0.5 \mu\text{rad}$ at 370 m and less than $0.01 \mu\text{rad}$ at 1,500 m implies at least inverse cube attenuation with distance. This may be expected from either a small scale or shallow source or a slightly decreased rigidity in or near the fault zone. The latter possibility seems most likely since the surface displacements extend at least 1.2 km along the fault trace for the Melendy observations and seem in other cases to extend to several tens of kilometres⁶. This would also explain why a search for inter-relationship between short period tilt changes on adjacent tiltmeters has so far been inconclusive.

The tilt signals for all three events are similar and therefore indicate a similar source process. The residual west tilt observed at MEL for the creep events, and the general systematic later arrivals and lower amplitudes of these and other creep events between MRC and XMR^{8,9}, indicate that the slip amplitude and/or the propagation velocity probably vary as the events propagate. Even with decreasing propagation velocity and slip amplitude to the north, however, the form of the tilt and the nearly simultaneous onset times for tilt and creep during the October 3 event (Fig. 2b) are difficult to fit with a single simple model of horizontal slip propagating northwards. The introduction of a vertical slip component (south-western side up) over a limited extent of the fault with a local maximum between MRB and MRC, together with horizontal slip, can generate time histories similar to those observed. The accumulated vertical displacement necessary as a consequence of this model is indicated by the presence of a north-easterly-facing scarp along the fault trace from MRW to MRC, with increasing scarp height to the north-west (the point of maximum development is about 3.5 m between MRB and MCR).

Most records of other creep events at Melendy Ranch display regular patterns of onset times^{8,9} although variations similar to those for the October 3 and April 14 events occur. Simple slip models, such as indicated above, are not always compatible with the detailed onset times and the creep amplitudes. It is possible that the exact onset times of creep, and perhaps also the amplitudes, result from, or are modified by, the particular local failure conditions of the surface material. More examples on a supplemented tilt and strain array in this area should clarify many of these details and make a rigorous inversion possible.

We thank Bill Stuart and Wayne Thatcher for critically reading the manuscript. A more complete discussion will be held elsewhere.

M. J. S. JOHNSTON
S. MCHUGH
R. O. BURFORD

National Center for Earthquake Research,
US Geological Survey,
Menlo Park, California 94025

Received October 27, 1975; accepted March 3, 1976.

- ¹ Johnston, M. J. S., and Mortensen, C. E., *Science*, **186**, 1031–1034 (1974).
- ² Steinbrugge, K. V., and Zacher, E. G., *Bull. seism. Soc. Am.*, **50**, 389–395 (1960).
- ³ Tocher, D., *Bull. seism. Soc. Am.*, **50**, 396–403 (1960).
- ⁴ Nason, R. D., *Earthq. Res. in E.S.S.A., E.S.S.A. Tech. Rep. E.R.L. 182—E.S.R.L. 11*, 12–15 (1970).
- ⁵ Nason, R. D., and Burford, R. O., *Earthq. Not.*, **43**, 25–28 (1972).
- ⁶ King, C.-Y., Nason, R. D., and Tocher, D., *Phil. Trans. R. Soc.*, **A274**, 355–360 (1973).

- ⁷ Burford, R. O., Allen, S. S., Lamson, R. J., and Goodreau, D. D., in *Tectonic Problems of the San Andreas Fault System*: Stanford Univ. Publ. geol. Sci., **13**, 268–274 (1973).
- ⁸ Yamashita, P. A., and Burford, R. O., *Catalogue of preliminary results from an 18-station creepmeter network along the San Andreas Fault System in Central California for the time interval June 1969–June 1973* (United States Geological Survey, National Center for Earthquake Research, Menlo Park, California, 1973).
- ⁹ Nason, R. D., Philippsborn, F. R., and Yamashita, P. A., *Catalogue of creepmeter measurements in central California from 1968 to 1972* (United States Geological Survey, National Center for Earthquake Research, Menlo Park, California, 1974).
- ¹⁰ Savage, J. C., Prescott, W. H., and Kinoshita, W. T., in *Tectonic Problems of the San Andreas Fault System*: Stanford Univ. Publ. geol. Sci., **13**, 44–53 (1973).
- ¹¹ Savage, J. C., and Burford, R. O., *J. geophys. Res.*, **78**, 832–845 (1973).

Compression and crustal shortening in Andean-type orogenesis

CONTROVERSY exists as to the role of crustal compression and shortening in the formation of the easterly-directed thrust system in eastern portions of the North American Cordilleran orogen^{1,2}. In fact, the controversy can be extended to the general issue of the presence or absence of crustal shortening in any major phase of the development of that orogen. Opponents of crustal shortening have suggested that most fold and thrust structures of the Cordilleran orogen can be explained by gravity-induced stresses^{3,4}.

Tectonic analogies between the modern central Andes and the US Cordillera of Mesozoic and earliest Tertiary age have been discussed by Hamilton^{5,6}, James⁷, and others. The Andes, a chain characterised by voluminous magmatic activity, are bounded on the west by the Peru–Chile trench, a zone of ongoing plate convergence, and on the east by an active fold and thrust belt which separates orogen from craton. Rates of subduction are of the order of 7–12 cm yr⁻¹ (ref. 8), and can reasonably be extrapolated back at least into the late Miocene on the basis of marine magnetic anomaly studies⁹. Since Middle Miocene time the central Andes, including Peru, have experienced intensive igneous activity and accompanying orogenesis¹⁰. Folds, reverse faults and thrust faults of eastward vergence have developed along the eastern margin of the Andes in the Sub-Andean zone (Fig. 1). The Andes were lifted up to their present height during this late Cainozoic episode of orogenesis.

Stauder¹¹ in a study of seismicity in the Peruvian Andes, has concluded that the Nazca Plate is thrusting under the South American plate beneath central and northern Peru along a surface of shallow (10–15°) dip. Of particular importance to an understanding of Andean orogenesis is Stauder's additional conclusion that the leading (western) edge of the South American plate is under east–west horizontal compression for a distance of approximately 700 km east of the Peru Trench (Fig. 1). Numerous hypocentres are present in that plate at depths of 1–90 km and for distances up to 700 km east of the trench. Most lie at distances from the trench of 200 to 600 km, depending on latitude. Focal mechanisms for 9 or 10 of these intraplate earthquakes yield horizontal compressive stress axes oriented approximately east–west. Three solutions indicate predominantly strike-slip faulting and six indicate reverse dip-slip faulting.

Middle Miocene to Recent central Andean structural and magmatic features can best be explained by the effects of convergence of the Nazca oceanic plate toward and beneath the South American Plate. Magma genetically related to subduction of the Nazca Plate has intruded the leading edge of the South American lithosphere and led to its thermal weakening. Deformation in response to east–west horizontal compression is localised in the plate margin region of elevated temperature where the lithosphere yields variably by fracture and flow. Compressive thickening of the leading edge can account for the uplift of the Andean Chain. The Sub-Andean fold and thrust belt thus marks the present eastward limit of intraplate deformation resulting from subduction of the Nazca Plate. Gravity may have

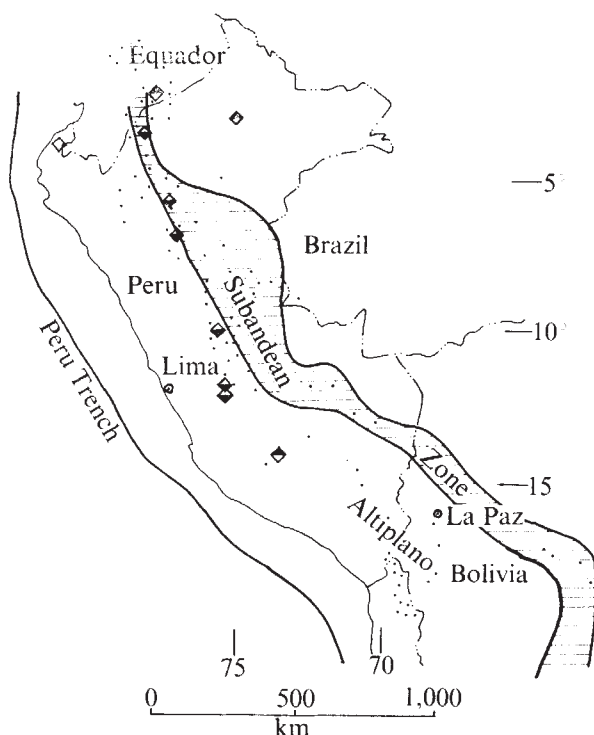


Fig. 1 Epicentres of shallow focus (1–70 km) earthquakes in the central and eastern part of the Peruvian sector of the Andes for 1961–68. Diamonds, earthquakes analysed by Stauder¹¹; diamonds with blackened upper half, strike-slip solution; diamonds with blackened lower half, reverse fault solution; open diamonds, solution for north–south compressive stress and reverse faulting.

a minor role in the tectonic development of the Andean Orogen, but it is secondary relative to lithospheric compression and shortening.

The geological history of the Mesozoic and early Cretaceous Cordilleran orogenic belt of western North America suggests it developed in a structural setting much like that of the modern Andes. Key questions about the origin of the Cordilleran belt concern the role of horizontal compression in crustal deformation, particularly in the formation of the foreland or marginal fold and thrust belt which, before Tertiary basin-range extension, lay up to 800 km east of the convergent plate margin. Stauder¹¹ has concluded that the leading edge of the South American Plate beneath the Andes is under horizontal east–west compression for a distance at least 700 km east of the Peru Trench. His data demonstrate that crustal compression and shortening are possible in an Andean structural setting far from the offshore plate boundary, and support a compressive origin for older Cordilleran structures analogous to those of the central Andes.

This work was supported by the NSF.

B. C. BURCHFIELD

Department of Geology,
Rice University,
Houston, Texas 77001

GREGORY A. DAVIS

Department of Geological Sciences,
University of Southern California,
Los Angeles, California 90007

Received August 6, 1975; accepted March 12, 1976.

¹ Burchfiel, B. C., and Davis, G. A., *Am. J. Sci.*, **272**, 97–118 (1972).

² Burchfiel, B. C., and Davis, G. A., *Am. J. Sci.*, **275-A**, 363–396 (1975).

³ Roberts, R. J., *Bull. geol. Soc. Am.*, **83**, 1989–2004 (1972).

⁴ Price, R. A., and Mountjoy, E. W., *Spec. Pap. geol. Assoc. Can.*, **6**, 7–25 (1970).

⁵ Hamilton, W., *Bull. geol. Soc. Am.*, **80**, 2409–2430 (1969).

⁶ Hamilton, W., *Bull. Ore. St. Dep. geol. Miner. Ind.*, **65**, 175–184 (1969).

⁷ James, D. E., *Bull. geol. Soc. Am.*, **82**, 3325–3346 (1971).

⁸ Plafker, G., *J. geophys. Res.*, **77**, 901–925 (1972).

⁹ Herron, E. M., *Bull. geol. Soc. Am.*, **83**, 1671–1692 (1972).

¹⁰ Noble, D. C., McKee, E. H., Farrar, E., and Petersen, U., *Earth planet. Sci. Lett.*, **21**, 213–218 (1974).

¹¹ Stauder, W. J., *J. geophys. Res.*, **80**, 1053–1064 (1975).

Caledonian magmatic activity in south-eastern Greenland

THE position of the Caledonian front in eastern Greenland is important^{1,2} in any reconstruction of the pre-drift North Atlantic (see ref. 3). No evidence of any Caledonian activity has yet been found south of Scoresby Sund, however, where the Caledonian Front, exposed in Gåseland⁴, is seen to pass under the thick plateau basalts of Kong Christian den IX's Land⁵. South of the basalts, only Precambrian ages have been obtained². We report here on a complex of Caledonian age in the Kangerdlugssuaq district, which lies about 300 km to the south of the Gåseland exposures. Unknown Caledonian events may be recorded in basement rocks of the areas, and may be uncovered by more detailed investigation. Moreover, petrological similarities between the new complex and the alkaline suite found in Assynt, north-western Scotland, may provide a further link in the North Atlantic jigsaw puzzle.

We have had access to specimens from the Batbjerg Intrusion at the head of Kangerdlugssuaq and have published a description of leucite from this locality, along with a map showing its location and approximate extent⁶. The rocks were believed to be Tertiary, in common with those of the nearby Gardiners Plateau Intrusion (which it resembles in many aspects)⁷ and other plutonic rocks of the area⁸. It was not until a visit in the summer of 1975 by a combined expedition from the Universities of Copenhagen and Toronto that many of the rocks in the Batbjerg Intrusion were found to be deformed and thus unlike the Tertiary rocks of the district. The impression that this complex is different from the nearby Tertiary rocks is confirmed by the new radiometric ages of around 440 Myr (Table 1). The dated rocks were selected as having primary magmatic textures, and we believe that the ages indicate crystallisation from a magma at the time of the Caledonian Orogeny.

In Scotland, alkaline rocks of a similar age⁸ occur in the Assynt district. In particular, the Loch Borolan laccolith⁹ is a layered complex of melanite pyroxenites, nepheline syenites and nordmarkitic syenites. Characteristic minerals are melanite garnet, acmitic diopside, green 'biotite', and rounded aggregates, interpreted as pseudoleucites.

Preliminary examination of the Batbjerg rocks shows that the bulk of the complex is composed of pyroxenites containing acmitic diopsides (an analysis has been presented in ref. 6), olivine, mica (sometimes green) and apatite. Other rocks include folded and sheared dunites, meltegitic rocks containing leucite and melanite, and a net-veined complex of syenitic dykes, some undersaturated, some granitic, cutting the pyroxenites. Many volcanic necks with associated cone sheets and ring dykes cut the surrounding gneisses, but they show extensive unroofing of the pyroxenites and may be foliated. Masses of metamorphosed limestone (some with crinoid fragments) and quartzites occur along the contact between the pyroxenites and the gneisses. They are interpreted as screens of country rock which have sunk from above.

It is clear that, superficially at least, there are many petrological similarities between Batbjerg and Loch Borolan. A speculative possibility is that the calcareous sediments at Batbjerg are equivalent to the Durness limestone and quartzites at Assynt, which had such a prominent role in Shand's now unpopular hypothesis for the development of the undersaturated rocks, both at Borolan and in

Table 1 Radiometric and petrographic data for the Batbjerg Complex, Kangerdlugssuaq

a Rb/Sr data			
Sample	$^{87}\text{Rb}/^{86}\text{Sr}^*$	$^{87}\text{Sr}/^{86}\text{Sr}^\dagger$	
NM 3206/5 Diopside	0.00303 ± 0.00002	0.7046 ± 0.0003	(unspiked)
NM 3209 Phlogopite	28.4 ± 0.2	0.8819 ± 0.0006	(spiked)
NM 3206/5 Phlogopite	56.6 ± 0.4	1.0494 ± 0.0006	(spiked)
b K/Ar data			
NM 3209 Phlogopite	^{40}Ar rad (standard $\text{cm}^3 \text{g}^{-1} \times 10^{-4}$)	1.4712	(mean of 2 determinations)
	^{40}Ar rad (%)	98.4	(mean of 2 determinations)
	K_2O (%)	9.54	(mean of 6 determinations by atomic absorption)
	Age (Myr)	446 ± 13	
c Petrographic notes			
NM 3209 Dark pegmatitic pyroxenite with abundant brown phlogopite			
NM 3206/5 Bright green, friable pyroxenite with about 10% pale greenish phlogopite			
Both rocks are free from deformational fabrics and felsic minerals.			

* Errors on $^{87}\text{Rb}/^{86}\text{Sr}$ ratios determined from 10 duplicates.

† Errors on $^{87}\text{Sr}/^{86}\text{Sr}$ ratios determined from 19 unspiked duplicates and 10 spiked duplicates.

Using $\lambda_{\text{Rb}} = 1.39 \times 10^{-11}$ the York-1 model provides an age of 443 ± 14 Myr with an initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of 0.7046 ± 0.0016 (ref. 2). Standard errors include a 'Student's t ' multiplier based on results for 10 duplicates.

other parts of the world¹⁰. Detailed petrological and mineralogical studies are at present being carried out on the Batbjerg rocks and will be published elsewhere. They should allow a more reliable comparison between the two intrusions. The gneisses surrounding the Batbjerg Intrusion are cut by an immense number of volcanic vents and associated dykes and cone sheets. The vents range in size from a few metres to perhaps half a kilometre in diameter and represent a further point of similarity to the Scottish and Irish Caledonides, where similar volcanic bodies, assigned to the appinite suite, are widespread^{11,12}. Like the appinites, the Batbjerg vents are rich in hornblende and were emplaced by highly explosive, volatile-rich magmas.

An important question to be asked is, how widespread is Caledonian activity in the Kangerdlugssuaq district and further to the south? Existing radiometric ages from the gneissic basement of the area, which include K/Ar, $^{40}\text{Ar}/^{39}\text{Ar}$, Rb/Sr mineral and whole rock ages^{2,13,14}, give ages ranging from ~ 1,700 to 2,700 Myr. Sampling has not, however, been carried out with radiometric dating in mind, and no systematic mapping of the basement has yet been made. It is clear that a great deal more detailed work of this nature must be undertaken before the history of the basement can be unravelled with any certainty. In recent years it has become clear that old ages frequently occur in younger fold belts as evidenced by, for example, the 3,000 Myr isochron reported for samples from within the Caledonides of the Scoresby Sund district¹⁵.

It is tempting to speculate that the close spatial association of similar rock types with widely differing ages is not coincidental. Thus, the Batbjerg Intrusion was originally regarded as Tertiary⁸ largely because of its petrological similarity to the Gardiners Plateau Intrusion which cuts Tertiary plateau basalts and is not more than 10 km distant. The Tertiary age of the Gardiner Intrusion is confirmed by an unpublished fission track age of 54.8 ± 3.4 Myr (A. G. W. Gleadow, personal communication). It comprises ultramafic rocks (dunites, pyroxenites and melilite rocks or uncomphagrites) with smaller amounts of nepheline syenites, urtites and related rocks. As such it falls into the general category of alkaline ultramafic complexes, to which the Batbjerg Intrusion is also assigned.

A possibility which should be explored, probably by means of isotopic studies, is that the origins of both intrusions are attributable to an alkali-enriched upper mantle which has underlain this area for several hundreds of millions of years, a possibility that has been suggested in other cases. Thus, it is now becoming clear that many

ultramafic nodules in lavas apparently preserve an isotopic record of earlier mantle events. For example, Burwell¹⁶ has recently reported nodules from Australia which lie on an isochron of about 700 Myr. That date coincides with a period of plate convergence similar to that in the North Atlantic during Caledonian times.

Until this report of Caledonian rocks in the Kangerdlugssuaq district, there has been no basis whatsoever for showing the Caledonian front in eastern Greenland passing out into the Denmark Strait near the mouth of Kangerdlugssuaq Fjord, as has been done by Flinn³. Indeed, rocks from the only basement exposure to the east of this line had already been reported to have radiometric ages of over 2,000 Myr. In the light of these considerations, it must be concluded that the position of the Caledonian front in this area is unknown and cannot therefore be used for pre-drift reconstructions. The discovery of rocks of Caledonian age in the Kangerdlugssuaq district indicates, however, that preconceived ideas on the basement rocks of southern Greenland are not necessarily correct. It is clearly desirable to carry out further work on the basement in that region.

We thank Ole Larsen and D.C. Rex for age determinations (Table 1). The field work was supported by the Danish Natural Sciences Research Council and the National Research Council of Canada. Preliminary samples were put at our disposal by the Northern Mining Company, Copenhagen.

J. D. Friderichsen and A. K. Higgins of the Geological Survey of Greenland have offered valuable comments.

C. K. BROOKS

*Institut for Petrologi,
Øster Voldgade 5,
DK-1350 Copenhagen K,
Denmark*

J. J. FAWCETT
J. GITTINS

*Department of Geology,
University of Toronto,
Toronto M5S 1A1,
Canada*

Received February 2; accepted March 10, 1976.

- 1 Wager, L. R., *Meddr Grønland*, **134**, 5 (1947).
- 2 Wager, L. R., and Hamilton, E. I., *Nature*, **204**, 1079 (1964).
- 3 Flinn, D., *Proc. Geol. Ass.*, **82**, 469 (1971).
- 4 Wenk, E., *Meddr Grønland*, **168**, 1 (1961).
- 5 Brooks, C. K., *Mem. Am. Ass. Petrol. Geol.*, **19**, 150 (1973).
- 6 Brooks, C. K., Ronsbo, J. G., and Nielsen, T. F. D., *Bull. geol. Soc. Danm.*, **24**, 93 (1975).
- 7 Frisch, W., and Keusen, H., *Schweiz. Min. Pet. Mitt.*, **55**, 132 (1975).

- ⁸ Brown, P. E., Miller, J. A., and Grasty, R. C., *Proc. Yorks. geol. Soc.*, **36**, 25 (1968).
⁹ Shand, S. J., *J. Geol.*, **47**, 408 (1939).
¹⁰ Shand, S. J., *Eruptive Rocks* (Wiley, New York, 1950).
¹¹ Mercer, E. P. C., *The Geology of Scotland* (edit. by Craig, G. Y.), 229 (Oliver and Boyd, Edinburgh, 1965).
¹² Pitcher, W. S., and Berger, A. R., *The Geology of Donegal* (Wiley-Interscience, New York, 1972).
¹³ Turner, G., in *Palaeogeophysics* (edit. by Runcorn, S. K.), 491 (Academic, London and New York, 1970).
¹⁴ Pankhurst, R. J., Beckinsale, R. D., and Brooks, C. K., *Contr. Miner. Petrol.* (in the press).
¹⁵ Rex, D. C., and Gledhill, A., *Bull. geol. Soc. Danm.*, **23**, 49 (1974).
¹⁶ Burwell, A. D. M., *Earth planet. Sci. Lett.*, **28**, 69 (1975).

Na, P, Ti and coordination of Si in garnet from peridotite and eclogite xenoliths

THEORY and experiment show that some cations are surrounded by extra oxygen anions at elevated pressure: for example, Si changes from 4-coordination in quartz and coesite to 6-coordination in stishovite. Sobolev and Lavrent'ev¹ suggested that substitution of Na in garnet derived from the mantle is coupled with 6-coordinated Si according to the equation $\text{Ca}^{\text{VIII}} + \text{Al}^{\text{VI}} \rightarrow \text{Na}^{\text{VIII}} + \text{Si}^{\text{VI}}$. We show here from sensitive electron-microprobe analyses that Na in garnets from ultramafic and eclogite xenoliths can be explained by coupled substitutions with Ti and P without invoking 6-coordinated Si.

Table 1 records mean values for simultaneous electron-microprobe analyses of selected garnets using a strong electron beam and long counting times. Backgrounds were carefully estimated at several wavelengths above and below the peak, and the standard deviation of individual analyses is 15–25 p.p.m. Cracks and surface blemishes were avoided during analysis with a 70- μm diameter spot, and 20 out of 21 garnets were found to be homogeneous from analyses at 3–6 places. Standards were $\text{Di}_{85}\text{Jd}_{15}$ glass (Na), A408 apatite (P) and Corning V glass (Ti). The correction factors for Na and Ti were trivial and were ignored, but correction factors for P were applied from the GLAB program (L. W. Finger, personal communication) using major-element concentrations from solid state detector and spectrometer analyses with a 2- μm diameter electron beam.

The concentration ranges for Na_2O (100–1,400 p.p.m. by weight) and TiO_2 (210–7,800) do not exceed those given previously for Na in garnets associated with diamond¹ and Ti in garnets from peridotite and eclogite xenoliths². The range for P_2O_5 (160–940) is within that for garnets synthesised from basaltic melts at 1.8–4.5 GPa (18–45 kbar) pressure³. No

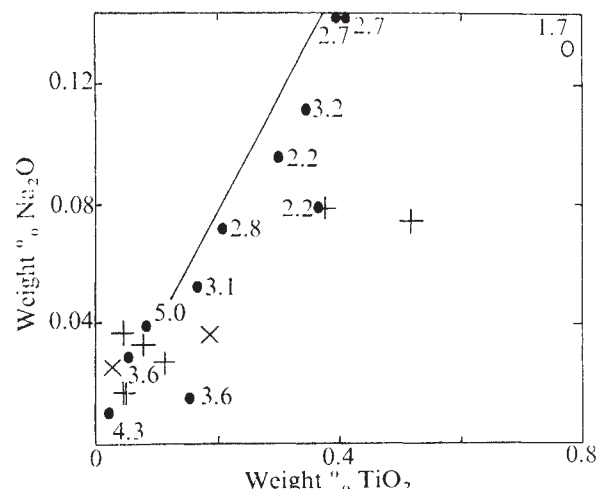


Fig. 1 Weight % Na_2O against TiO_2 in mantle-derived garnets: +, garnet-lherzolite; ×, other ultramafic rock; ●, eclogite; ○, diopside megacryst with garnet and sulphide inclusions. The numbers are K_D as defined in Table 1 for the eclogites. The straight line represents equal atomic concentrations of Na and Ti.

previous analyses of P_2O_5 in mantle-derived garnets have been reported.

The Na concentration generally increases with the Ti concentration (Fig. 1), and in all but six garnets, atomic Ti exceeds atomic Na. A rough correlation exists also between Na and P, but the spread is well outside the analytical error band. In all but one garnet atomic Na exceeds atomic P.

Because atomic Na is always less than atomic (Ti+P), there is no need to invoke 6-coordinated Si (Fig. 2). Substitutions of the type $\text{Ca}^{\text{VIII}} + \text{Si}^{\text{IV}} \rightarrow \text{Na}^{\text{VIII}} + \text{P}^{\text{IV}}$ and $\text{Ca}^{\text{VIII}} + \text{Al}^{\text{VI}} \rightarrow \text{Na}^{\text{VIII}} + \text{Ti}^{\text{VI}}$ retain the normal valence and coordination number of Si, Ti and P. Thilo⁴ reported garnet structure for $\text{Na}_3\text{Al}_2\text{P}_3\text{O}_{12}$ and $\text{NaCa}_2\text{Mg}_2\text{P}_3\text{O}_{12}$ compositions. Ringwood and Lovering⁵ suggested that $(\text{CaNa}_2)\text{Ti}_2\text{Si}_3\text{O}_{12}$ and $(\text{Ca}_2\text{Na})(\text{AlTi})\text{Si}_3\text{O}_{12}$ are the probable Na-bearing components in garnet synthesised from a clinopyroxene-ilmenite nodule at pressure > 10 GPa. Although 6-coordinated Si may occur in the lower mantle as stishovite (see ref. 6), there is no evidence for its existence in minerals ejected to the Earth's surface from the upper mantle.

Table 1 Analyses of mantle-derived garnet

Label	Host rock*	Location	Weight			K_D
			TiO_2 (p.p.m.)	P_2O_5 (p.p.m.)	Na_2O (p.p.m.)	
740	garnet-lherzolite	Lashaine	460	200	170	2.1
794	garnet-lherzolite	Lashaine	430	320	170	2.4
749	garnet-wehrlite	Lashaine	1,900	300	380	2.3
1143B	garnet-lherzolite	Bultfontein	790	280	330	2.8
1149	garnet-lherzolite	Bultfontein	460	690	370	2.5
1156	garnet-lherzolite	Bultfontein	5,200	360	750	2.2
1359	garnet-lherzolite	Matsoku	3,800	500	790	2.2
1755	garnet-pyroxenite	Klipfontein	290	390	250	3.6
1870/2	garnet-lherzolite	Letseng-le-Terai	1,130	300	270	3.4
1178	eclogite	Roberts Victor	4,100	500	1,400	2.7
1187	eclogite	Roberts Victor	3,000	530	950	2.2
1188	eclogite	Roberts Victor	820	230	390	5.0
1190	eclogite	Roberts Victor	3,400	800	1,120	3.2
1191	eclogite	Roberts Victor	1,700	250	520	3.1
1692	eclogite	Roberts Victor	4,000	520	1,400	2.7
1780/1	eclogite	Newlands	3,600	710	790	2.2
1780/2	eclogite	Newlands	2,100	460	720	2.8
1829/1	eclogite	Bobbejaan	550	910	280	3.6
1829/2	eclogite	Bobbejaan	210	160	100	4.3
1934	eclogite	Vissuri	1,600	260	150	3.6
73-267	clinopyroxene megacryst	Monastery	7,800	940	1,300	1.7

$K_D = (\text{FeO}/\text{MgO})_{\text{gt}}/(\text{FeO}/\text{MgO})_{\text{cpx}}$.

* All garnet lherzolite, pyroxenite and wehrlite host rocks exhibit granular texture.

The validity of geothermometers and geobarometers involving pyroxene and garnet is controversial, but the parameter $K_D = (\text{FeO}/\text{MgO})_{\text{gt}}/(\text{FeO}/\text{MgO})_{\text{cpx}}$ has been used to classify eclogites on a probable depth basis (see ref. 7), and has been shown from laboratory syntheses⁸ to decrease with increasing temperature and decreasing pressure. There is a weak tendency for Na, Ti and P in the garnets from eclogites to increase as K_D decreases (for example in Fig. 1), which might

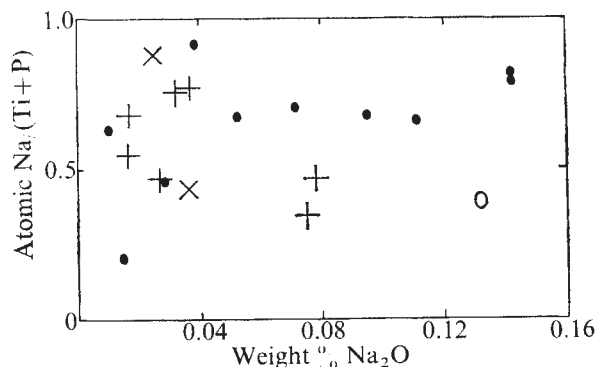


Fig. 2 Weight % Na_2O against atomic $\text{Na}/(\text{Ti}+\text{P})$. See Fig. 1 for symbols.

suggest that substitution of these minor elements in garnet increases with the depth of formation. Elsewhere we show that the correlation between K_D and $\text{Na}_{\text{gt}}/\text{Na}_{\text{cpx}}$ for lherzolites and eclogites is good. Much further work is needed on the factors which control the substitution of minor and trace elements in mantle minerals, especially because partial melting should lead to strong fractionation.

We thank O. Draughn, T. Solberg and I. M. Steele for help and the NSF and Materials Research Laboratory for support.

F. C. BISHOP

J. V. SMITH

Department of the Geophysical Sciences,
University of Chicago,
Chicago, Illinois 60637

Department of Geology,
University of St Andrews,
St Andrews, UK

J. B. DAWSON

Received March 4; accepted March 18, 1976.

¹ Sobolev, N. V., Jr. and Lavrent'ev, Ju. G., *Contr. Miner. Petrol.*, **31**, 1–12 (1971).

² Khar'kiv, A. D., and Makovskaya, N. S., *Dokl. Akad. Nauk SSSR*, **193**, 124–126 (1969).

³ Thompson, R. N., *Earth planet. Sci. Lett.*, **26**, 417–424 (1975).

⁴ Thilo, E., *Naturwissenschaften*, **29**, 239 (1949).

⁵ Ringwood, A. E., and Lovering, J. F., *Earth planet. Sci. Lett.*, **7**, 371–375 (1970).

⁶ Liu, L., *Nature*, **258**, 510–512 (1975).

⁷ Banno, S., *Phys. Earth planet. Interiors*, **3**, 405–421 (1970).

⁸ Råheim, A., and Green, D. H., *Contr. Miner. Petrol.*, **48**, 179–203 (1974).

Tree growth and climatic cycles in the rain shadow of the Grampian mountains

TREE-RING widths have been widely studied as a guide to historical meteorology^{1,2}, the assumption being that fluctuations superimposed on the smooth progression of change with age are climate induced. Several authors^{3,4} have even identified cyclical patterns in tree growth, some of which they have tentatively linked to sunspot activity,

Table 1 Correlations between ring width deviation, May rain during the same year, annual temperature during the same year and rain during May plus June of the previous year

	<i>D</i>	<i>M</i> ^{0.1}	<i>T</i>	<i>MJ</i>
<i>D</i>	—	0.453	0.379	0.423
<i>M</i> ^{0.1}	<i>P</i> < 0.01	—	−0.095	0.176
<i>T</i>	<i>P</i> < 0.05	NS	—	0.220
<i>MJ</i>	<i>P</i> < 0.01	NS	NS	—

but generally the precise climatic parameter that may be operating remains unspecified. In Scandinavia, ring widths seem to fluctuate with summer temperature^{5,6}, although precipitation during early summer may also be influential⁷. In Scotland, on the other hand, Schove and Frewer⁸ were unable to relate any meteorological character to the growth pattern of native Scots pine (*Pinus sylvestris* L.) on Speyside. Here we report, however, that examination of the results from two long term fertiliser experiments in mature Scots pine at Alltcaileach forest (National Grid Reference NO 3596), on upper Deeside, has shown that growth exhibits several regular oscillations each of which can be related to a particular climatic cycle. Furthermore, it would seem that some of these climatic cycles may be limited to low rainfall areas, such as the rain shadow of the Grampian mountains.

The trees, planted in 1883 on a freely drained humus iron podzol developed on a terrace of the river Dee, remained virtually unthinned until 1934. Since then numerous light thinnings have been carried out at no fixed interval. In late 1972 increment cores were drilled from opposite sides of 32 untreated trees and the width of each of the 40 outermost rings measured to the nearest 0.01 mm. Meteorological data used is that published⁹ for Balmoral (~10 km due west and upstream) where the mean annual rainfall and daily temperature are 830 mm and 6.5 °C, respectively.

During the 40 yr, 1933–1972, average ring widths declined linearly ($P < 0.05$) and the deviations have been taken from the fitted line. Preliminary inspection of a segment of the data suggested that correlations with winter climate were very unlikely. Ring width deviations, therefore, were tested for correlations with rainfall amounts and mean temperature recorded over the full year and during the six summer months April–September, individually and in all possible combinations. Only three climatic parameters were significantly and independently related to ring width deviations (Table 1), two of which—annual temperature and the previous year's May–June rainfall—were linearly related, whereas the positive response per unit of May rainfall became less pronounced as the amount increased. These three terms are all significant when included in the following multiple regression, the multiple correlation coefficient for which is 0.67: $D = -2.8765 + (1.2339 \pm 0.3555)M^{0.1} + (0.1371 \pm 0.0485)T + (0.0012 \pm 0.0006)MJ$ where *D* is the ring width deviation (mm), *M* is May rain of the same year (mm; *M* 62.5, s.d. 30.4), *T* is annual temperature of the same year (°C; *T* 6.48, s.d. 0.55) and *MJ* is the total rain during May–June of the previous year (mm; *MJ* 113.1, s.d. 45.0).

Ring width deviations and these three climatic parameters were then subjected to spectral analysis followed by repeated curve fitting about the indicated harmonics. Three significant oscillations, and one just short of significant, were found in

Table 2 Climatic oscillations associated with the oscillations in ring width deviation shown in Fig. 1

	Period (yr)	Significance of fit	Amplitude ±	Mean	Lag between climate and ring width (yr)
Annual temperature	42.0	<i>P</i> < 0.01	0.34 °C	6.48 °C	None
May rain	23.0	<i>P</i> < 0.01	17.9 mm	62.5 mm	0.6
May rain	11.9	<i>P</i> < 0.01	21.0 mm	62.5 mm	None
June rain	4.44	<i>P</i> < 0.05	15.1 mm	50.9 mm	1.2

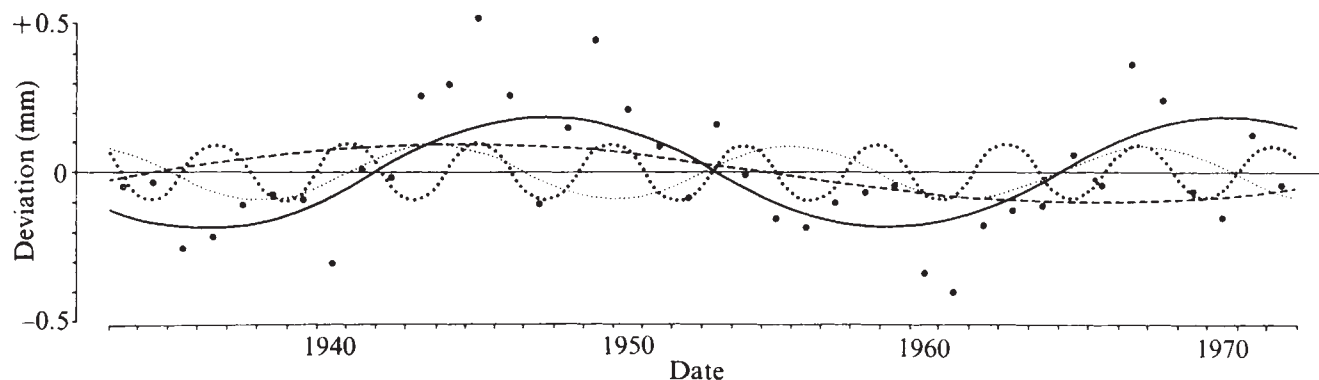


Fig. 1 Observed values of ring width deviation (●) and fitted oscillations; ---- 42.0-yr period with amplitude 0.094 mm ($P < 0.05$), — 23.0-yr period with amplitude 0.184 mm ($P < 0.001$), 4.44-yr period with amplitude 0.093 mm ($P < 0.05$) and a barely significant oscillation (fine dotted line) of 11.9 yr with amplitude 0.091 mm ($P < 0.1$).

Table 3 Occurrence of the four climatic oscillations associated with tree growth at Alltailleach forest at meteorological stations both to the west and east of the Grampian mountains

Station	Latitude	Longitude	Parameter and period (yr)			
			Annual temperature 42.0	23.0	May rain 11.9	June rain 4.44
Western						
Stornoway	58°13'N	06°20'W	$P < 0.01$	NS	NS	NS
Rothsay	55°50'N	05°04'W	NS	NS	NS	NS
Fort Augustus	57°08'N	04°40'W	$P < 0.05$	NS	NS	NS
Eastern						
Nairn	57°35'N	03°54'W	$P < 0.05$	NS	$P < 0.05$	NS
Braemar	57°00'N	03°24'W	NS	$P < 0.001$	$P < 0.05$	($P < 0.1$)
Balmoral	57°02'N	03°12'W	$P < 0.01$	$P < 0.01$	$P < 0.01$	$P < 0.05$
Aberdeen	57°08'N	02°08'W	$P < 0.05$	$P < 0.01$	$P < 0.001$	NS

the ring width deviations (Fig. 1); for each there was a matching oscillation in climate (Table 2) and there were no significant oscillations in these climatic parameters that were not reflected in the ring width deviations.

The absence of any previous mention of these relationships with tree growth in the British Isles, and the predominant influence of spring drought, promoted the idea that the climatic oscillations in Table 2 may be less pronounced, or may not occur, outside the rain shadow of the Grampian mountains. Accordingly the same climatic oscillations were looked for in three stations to the south and west and four stations to the north and east of the mountains (Table 3), the prevailing winds being south-westerly. The effect of the mountains on the oscillations in rainfall is quite striking, indeed the short term oscillation in June rain could only be detected at Balmoral and, less certainly, Braemar, two stations well up the Dee valley and tucked in behind the main mountain mass. These oscillations, therefore, may reflect variations in the strength of the westerlies. Further support for this suggestion can be found in a strikingly close similarity between the ring width pattern found at Alltailleach and patterns published recently¹⁰ for pine growing within the rain shadow of the Norwegian coastal mountains.

Experimental interest in this tree stand was originally prompted by the belief of local Forestry Commission management staff that tree health and growth were declining severely. It now seems that the decline from a peak positive deviation in 1945 to a large negative deviation in 1960 (Fig. 1) is primarily a reflection of the 23-yr cycle in May rain with the addition of the longer term temperature cycle. Irrespective of the experimental fertiliser application given in 1960, tree growth improved over much of the next decade but as the temperature was at the bottom of its cycle the optimum conditions of 1945 did not recur.

The number and length of these climatic cycles, coupled with the geographical variability in both their occurrence and their significance for tree growth, must counsel caution over recent attempts to explain patterns and trends of tree growth in terms of atmospheric pollution.

HUGH G. MILLER
JEAN M. COOPER

The Macaulay Institute for Soil Research,
Craigiebuckler,
Aberdeen AB9 2QJ, UK

Received January 22, accepted March 3, 1976.

- Bray, J. R., *Adv. ecol. Res.*, 7, 177-233 (1971).
- Glock, W. S., *Bot. Rev.*, 7, 649-713 (1941).
- Douglas, A. E., *Climatic cycles and tree growth, III. A study of cycles* (Carnegie Institution, Washington, 1936).
- Hasek, J., *Sb. vys. Šk. Zemed. Brno*, C40, 125-137 (1971).
- Mikola, P., *Fennia*, 75, 69-76 (1952).
- Schove, D. J., *Geogr. Annl.*, 36, 40-80 (1954).
- Jonsson, B., *Rapp. och Uppsatser, Inst. for Skogsproduktion, Skogshogskolan*, 16 (1959).
- Schove, D. J., and Frewer, A., *Scott. For.*, 15, 63-71 (1961).
- Meteorological Office, *Monthly Weather Reports*, 50-89 (HMSO, London, 1933-72).
- Abrahamsen, G., Horntvedt, R., and Tveite, B., *Sur Nedbors Virkning pa Skog og Fisk, Fragrapport Fr 2/75* (1975).

Energy expenditure in small children of obese and non-obese parents

SEVERAL recent studies have suggested the possibility of identifying some of the factors in infancy which contribute to the development of obesity in later life. For example, the abnormally rapid gain in weight of those infants who tend to become overweight later may be associated either with some genetic factor or with early overfeeding^{1,2}. It is, however, highly probable that a child born to obese parents will itself become obese³. The main aim of this study was to see if it was

possible to identify a difference in energy expenditure between children of 'normal' and 'obese' parents.

The energy intake and expenditure have been estimated in two groups of 4–5-yr-old children distinguished, not on the basis of their own weight, but according to the classification of their parents as being obese or non-obese. We found a difference which could be the result of either genetic or environmental factors. At this early age none of the children was significantly overweight for age or height, and the interest of this approach lies in the fact that any such differences observed may relate to factors contributing to obesity in later life.

Intake and expenditure measurements were made on 17 children and in a further three expenditure only was estimated. Height, weight and skinfold thickness were measured in all subjects. Body density was calculated from the regression equation of Brook⁴ and body fat from the equation given by Siri⁵. Lean body mass (LBM) was obtained by subtracting the fat content from the total body weight. Apart from age and apparent good health, the main criterion of selection of subjects was the willingness of the parents and children to take part. The children were divided into two groups, those with one parent whose adult weight had at some time been more than 20% above desirable weight for height⁶ (group O), and the remainder, neither of whose parents had ever been overweight (group N).

Daily energy expenditure was estimated from the integrated pulse rate for the day, based on the relationship between oxygen consumption and pulse rate. To establish this relationship, or to 'calibrate' the children, simultaneous measurements of oxygen consumption and heart rate were made on each child while he or she was lying down at rest and again while pedalling a stationary tricycle. Oxygen consumption was measured with a continuous air flow diaferometer, thereby eliminating the need for a face mask. Wherever possible more than two points on the calibration line were obtained. Integrated heart rate was then recorded on a SAMI (Socially Acceptable Monitoring Instrument) for 4 to 7 days and nights while the child went about his or her normal activities.

Energy intake was measured by analysing duplicate portions of all the food consumed by each child over a period of 5–7 d and the energy content of these portions estimated by ballistic bomb calorimetry⁷. The consumption of drinks was recorded separately and the energy content calculated from food tables^{8,9}.

Table 1 shows that although the two groups do not differ in body size, the estimates of energy intake and expenditure, including expenditure at rest, are significantly lower in group

O than in group N. The degree of significance is greater when the results are expressed per kg body weight. The correlation coefficients of body weight, body fat and LBM with energy intake, expenditure and resting rate of expenditure have been calculated for all subjects. The only significant correlations are those of resting oxygen consumption with body weight ($r + 0.58$) and LBM ($r + 0.64$). The results show therefore that the resting oxygen consumption is affected by both body weight and by having one obese parent. Analysis of covariance shows that having an obese parent is highly significant over and above the effect of body weight, the latter tending if anything to mask the effect of the former.

The most critical questions arising from this study are whether the differences between the two groups result from metabolic or environmental factors or both, and whether these in turn are the result of genetic or acquired differences.

Energy is required for maintenance, physical activity and growth, the last accounting for only about 2% of the total expenditure¹⁰. The major portion is used for maintaining and resynthesising body tissue, to cover the heat increment of feeding and the minimum level of muscular activity needed for sustaining body functions. Energy expended while a subject is relaxed and inactive, as were the children in this study when they were at rest, would be expected to be similar to their maintenance requirement. The mean resting energy expenditure of all the children calculated on a 24 h basis was $1,110 \pm 190$ kcalorie.

The daily requirement of energy for maintenance has been estimated¹⁰ to be ~ 105 kcalorie per $\text{kg}^{0.75}$ or $1.5 \times$ basal metabolic rate. A similar figure was obtained by Spady¹¹ based on a study of energy balance in children recovering from malnutrition. For the children in this study 105 kcalorie per $\text{kg}^{0.75}$ averages 970 kcalorie. This is close to the average resting expenditure of the two groups (1,110 kcalorie). It seems reasonable, therefore, to take the difference between the observed total expenditure and resting expenditure as an estimate of the energy cost of voluntary physical activity. On this basis the energy expended on physical activity by the children in group O (190 kcalorie) was only half that expended by the children in group N (371 kcalorie).

It is of note that in a study carried out in the 1930s (ref. 12), the mean energy intake of 4–5-yr-old children was found to be 1,770 kcalorie. The maintenance requirement calculated according to the FAO/WHO estimate¹⁰ would have been about 970 kcalorie, leaving 800 kcalorie available for physical activity —2–4 times that of the children in the present study. We know of

Table 1 Mean daily energy intake and expenditure

Body size	Group N	Differences between groups†		Differene	
		s.d.	Group O		s.d.
Height (cm)	111.3	7.4	110.7	7.5	*
Weight (kg)	19.08	3.53	19.49	3.46	*
Body fat (%)	14.3	5.2	16.8	3.7	*
LBM (kg)	16.2	2.24	17.0	1.71	*
Standard weight for height (%)	99.9	9.5	100.8	6.0	*
Energy measurements (kcalorie)					
Daily intake	1,433	188	1,115	300	$P < 0.01†$
Intake per kg body weight	75	13	57	8	$P < 0.01†$
Daily expenditure	1,508	352	1,174	297	$P < 0.05†$
Expenditure per kg body weight	79	18	60	9	$P < 0.02†$
Daily expenditure at rest	1,183	184	999	146	$P < 0.05†$
Expenditure at rest per kg body weight	62	9	52	6	$P < 0.02†$
Daily expenditure available for activity	371	306	190	173	$P < 0.2^*$
Groups combined					
Daily intake	1,325	263			
Daily expenditure	1,375	364			
Daily expenditure at resting rate	1,110	190			
Daily expenditure available for activity	265	266			

* Not significant.

† Significant.

‡ Number in group N = 12; number in group O = 8 (except for intake study when it was 5).

no evidence to suggest that basal metabolic rates of children have fallen dramatically over this time, and the conclusion that the decrease in intake is associated with changes in patterns of behaviour seems inevitable. These, in turn, may reflect changes in the attitudes of parents towards expected levels of activity and play of children. Other aspects of living such as television and domestic heating standards may also be contributory.

We found a difference between the energy expenditure of children of normal and obese parents which is of the same order (350 kcalorie d^{-1}) as the decline in average intakes over the past 40 yr. Although this decline is likely to be due to environmental changes affecting voluntary activity levels, the children of obese parents also have lower resting rates of energy expenditure, and it will be of great importance in future studies to determine if this is the result of behavioural or of inborn metabolic differences.

This work was supported by the MRC, Rank Hovis McDougall Ltd, and the Weight Watchers.

MARY GRIFFITHS
PHILIP R. PAYNE

Department of Human Nutrition,
London School of Hygiene and Tropical Medicine,
Keppel Street,
London WC1E 7HT, UK

Received December 8, 1975; accepted March 3, 1976.

- ¹ Eid, E. E., *Br. med. J.*, **2**, 74–76 (1970).
- ² Taitz, L. S., *Br. med. J.*, **1**, 315–316 (1971).
- ³ Seltzer, C. C., and Mayer, J., *Ann. N.Y. Acad. Sci.*, **134**, 688–695 (1965).
- ⁴ Brook, C. G. D., *Archs Dis. Childh.*, **46**, 182–184 (1971).
- ⁵ Siri, W. E., UCRL, 3349, *Dormer Laboratory*, University of California (1956).
- ⁶ Geigy, J. R., *Documenta Geigy*, Sci. tables, 172, (Basle, Switzerland, 1970).
- ⁷ Miller, D. S., and Payne, P. R., *Br. J. Nutr.*, **13**, 501–508 (1959).
- ⁸ McCance, R. A., and Widdowson, E. M., *MRC. Spec. Rep. Ser.*, No. 295, 270 pp. (HMSO, London, 1960).
- ⁹ *Sainsbury's Food Tables*, 28 pp. (J. Sainsbury Ltd, London 1973).
- ¹⁰ Joint FAO/WHO ad hoc, *Expert Committee Report on Energy and Protein Requirements*, FAO, Nutr. Migs Rep. Ser., No. 257, 118 pp (HMSO, London, 1973).
- ¹¹ Spady, D. W., *Am. J. clin. Nutr.* (in the press).
- ¹² Widdowson, E. M., *MRC. Spec. Rep. Ser.*, No. 257, 96 pp (HMSO, London, 1947).

Molecular evidence for dual origin of mangabeys among Old World monkeys

THE morphological similarities among the various mangabey species have been uniformly judged by primate systematists to warrant their association in a single genus (*Cercocebus*) of Old World monkeys. These similarities are not, however, reflected in their macromolecules. We find that the two groups of mangabeys are, by Old World monkey standards, so different at the protein level as to force the conclusion that the genus is not a natural that is, monophyletic, group.

Molecular data have become an important, independent source of phylogenetic information. Though their application to specific problems in systematics has been limited, they have revealed a common origin for the New and Old World higher primates^{1–3}; the monophyly of the pinnipeds among the artoid canivores (ref. 4 and paper presented at Society for Systematic Zoology meetings, 1975); the ursid nature of the giant panda^{5,6}, and the association of the tree frog *Hyla wrightorum* with *H. eximia* instead of *H. regilla*⁷. In each case, however the problems were longstanding and the possible solutions available; the molecular data provided independent and objective ways of choosing among the possibilities. On the other hand, new sources of information can point out unsuspected situations, such as the strong probability that man is as closely related to the chimpanzee and gorilla as they are to each other (refs 1, 8–10 and an unpublished result).

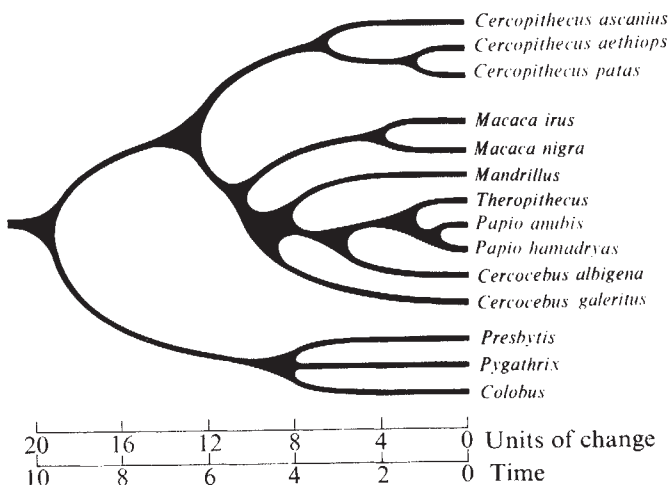
The mangabey situation seems to be an example of this latter phenomenon, involving the probable misplacement of two species (*C. albigena* and *C. aterrimus*) among the Old World monkeys. We present here evidence that the two mangabey

species groups—one containing *C. galeritus* and *C. torquatus* and the other *C. albigena* and *C. aterrimus*—may well be cladistically unrelated. We find our various sera representing the former species group (obtained as *C. fuliginosus* from the Davis Primate Center, and *C. galeritus* and *C. torquatus* from the Delta Primate Center) to be virtually indistinguishable both immunologically and electrophoretically. On the other hand, they are as distinct by these criteria from *C. albigena* and *C. aterrimus* as from any of the other 42-chromosome cercopithecine species (macaques and baboons). Finally, the *albigena-aterrimus* group seems to be an early offshoot of the baboon lineage, diverging from it some time after the separation of the *Mandrillus* (drill and mandrill) and *Macaca* lines and well before the *Papio-Theropithecus* radiation.

The Old World monkeys are divided into two subfamilies, Cercopithecinae and Colobinae, with the former containing the genera *Papio*, *Theropithecus*, *Cercocebus*, *Macaca* and *Cercopithecus*. These five contain some 35 species, and the strong evolutionary affinities among at least the first four are evidenced most strikingly by the ease with which viable intergeneric hybrids are produced^{11,12}. With so complex a grouping of closely related forms and the absence of a detailed fossil record¹³, we felt that a molecular study would prove useful in adding a new dimension to our systematic knowledge of the group, and the *albigena-aterrimus* finding is one striking result.

Antisera to numerous purified Old World monkey albumins and transferrins¹⁴ were prepared in rabbits (Dutch belted, three or four per immunogen) according to published procedures^{3,15} and pooled in reciprocal proportion to their micro-complement fixation (MC'F) titres. Differences among the various albumins and transferrins were measured using the MC'F procedure and are expressed as immunological distances³. Immunological distances are highly correlated with, and linearly proportional to, the number of amino acid sequence differences among the protein species being compared¹⁶. They are also additively apportionable along derived cladograms^{2,3,17,18} and thus serve as reliable indicators of phylogenetic affinities among the taxa being compared. The data reported here were obtained using antisera to the albumins of *Papio cynocephalus*, *P. papio*, *Macaca irus*, *Cercopithecus*

Fig. 1 Albumin plus transferrin phylogeny of the Old World monkeys. This cladogram has been derived using procedures detailed in ref. 2. The complete data sets will appear in a comprehensive article on Old World monkey systematics. Immunological distances among the albumins and transferrins of the various species were obtained using the antisera listed in the text. The immunological distance separating any two forms was apportioned additively along the two lineages using antisera directed to the proteins of species outside the taxon involved. All such lineage pairs were then superimposed along common segments while minimising the differences among the experimental immunological distances and those derived from the cladogram.



aethiops and *Presbytis entellus*; and to the transferrins of *Theropithecus gelada*, *P. cynocephalus*, *P. papio*, *P. hamadryas*, *M. irus*, *M. nigra*, *M. speciosa*, *C. torquatus*, *C. patas*, *P. entellus* and *Colobus guereza*.

The detailed molecular systematics of the Old World monkeys will be presented elsewhere, but it is necessary to place the major lineages relative to one another before considering the position of the *albigena-aterrimus* group among them. The phylogeny is presented in Fig. 1 and its reliability can be assessed by comparing the immunological distances calculated from it and the experimental values. With each of the latter is associated an uncertainty deriving from non-reciprocity and experimental error, and these have appreciable significance given the very small distances involved in this study. The non-reciprocity of the input data matrix is $\pm 1-2$ units and the experimental error of each measurement is also between 1 and 2 units. Given this level of uncertainty or noise in the input data, the mean difference of 2 units between the elements of the input and output data half-matrices can only be taken to indicate that we would find essentially perfect additivity if we could deal with a set of data uncontaminated by immunological noise. But the observed noise level of 2 units should be remembered when assessing the reliability of the lineage placements in Fig. 1. We have attempted to indicate this visually by making the nodes areas rather than points. We then find the 42-chromosome Cercopithecinae (Papionini of Jolly^{20,21}) forming a clade relative to *Cercopithecus*; within that clade *Theropithecus-Papio*, *Macaca*, *Mandrillus* and the *galeritus-torquatus* mangabeys form the major lineages.

The placement of the *albigena-aterrimus* lineage within this phylogenetic framework follows from a consideration of the data in Table 1 plus the albumin plus transferrin distances from *Papio* to: *Theropithecus*, 10 units; *Mandrillus*, 18; *Macaca*, 27; and *Cercopithecus*, 34. The distance between the two mangabey groups is 26 units, and the evolutionary significance of these observations is made evident when two colobines (*Presbytis* and *Colobus*) are used as reference species to assess the relative amounts of change in the two molecules along the lineages of interest. The colobine-*Papio* distance is 50 units; to the *galeritus-torquatus* mangabeys it is 51, and to the *albigena-aterrimus* mangabeys, 50. Thus in this case the albumins and transferrins have accumulated equal numbers of substitutions in the three lineages; therefore the fact that the two *albigena-aterrimus* molecules are more similar to their *Papio* counterparts than to those of the other mangabeys cannot be ascribed to a lack of change along either the *albigena-aterrimus* or *Papio* lineage. Thus we must conclude that the immunological data do not associate the two mangabey groups subsequent to the adaptive radiation of the Papionini and do indicate that the *albigena-aterrimus* lineage is probably an early offshoot of that leading to *Papio* and *Theropithecus*. This conclusion concerning the mangabeys is, of course, not the only interesting one from this study. The others will form subjects of future reports, but we note here the following. The various *Papio* species and *Theropithecus* are seen as very closely related at all levels surveyed here—albumin and transferrin immunology and plasma protein electrophoresis. Similarly the patas monkey, *Cercopithecus* (or *Erythrocebus*) *patas*, is found to be more closely related to the green monkey (*Cercopithecus aethiops*) than is any other *Cercopithecus* species. Finally the drill and mandrill (*Mandrillus*) form a lineage which is not specifically associated with *Papio-Theropithecus* (and, now, the *albigena-aterrimus* mangabeys).

In addition to the immunological studies just discussed, we have carried out some preliminary electrophoretic comparisons of the various sera used in this study. In the conditions used (7% acrylamide in a continuous 0.1 M Tris, 0.05 M glycine buffer), we typically see 20–25 serum protein bands staining with Coomassie blue. One can then use as a measure of similarity (*S*) between two samples simply the ratio of the number of bands with the same mobility in the two to the number of bands in the sample containing the smaller number.

Table 1 Albumin plus transferrin immunological distances between *Cercocebus* and other Old World monkeys

	<i>C. galeritus</i>	<i>C. albigena</i>
<i>C. galeritus</i> *	0	26
<i>C. albigena</i>	26	0
<i>T. gelada</i>	25	16
<i>Papio</i> †	24	12
<i>Mandrillus</i>	16	12
<i>Cercopithecus</i>	20	30
<i>Macaca</i> ‡	16	18
<i>Colobinae</i> §	51	50

*Antisera made to the albumin of a *C. galeritus* from the Delta Primate Center and to the transferrin of a *C. torquatus* from the Davis Primate Center fail to distinguish any of the various *galeritus-torquatus* individuals from each other.

†Using antisera to the albumins of *P. papio* and *P. cynocephalus* and to the transferrins of *P. papio*, *P. cynocephalus*, *P. anubis*, *P. hamadryas* and *P. ursinus*.

‡Using antisera to the albumin of *Macaca fascicularis* and to the transferrins of *M. fascicularis*, *M. nigra* and *M. speciosa*.

§Using antisera to the albumin of *Presbytis entellus* and to the transferrins of *P. entellus* and *Colobus guereza*.

For intergeneric comparisons involving *Macaca*, *Papio*, *Cercocebus galeritus*, and *Mandrillus*, the observed *S* values are in the range 0.1–0.2. The *S* value between *P. cynocephalus* and each of the following is: *P. hamadryas*, 0.75; *T. gelada*, 0.6; and *C. albigena* or *aterrimus*, 0.35. Among the various *torquatus-galeritus* individuals in our collection, the *S* values are no less than 0.8; for the *albigena-aterrimus* comparison the value is 0.8; and between individuals in the two mangabey groups it is 0.15–0.2. Thus these preliminary electrophoretic data are consistent with the immunological picture.

Previous electrophoretic studies have detected marked differences in the mobilities of the α and β chains of *C. albigena* haemoglobin as compared with those of *C. galeritus*²¹. In addition, the peptide pattern of *C. albigena* haemoglobin is more like that of *Papio* than of *Macaca*²². This then completes a phylogenetic congruence among all the available molecular data relevant to mangabey systematics. We suggest that a DNA hybridisation study would provide useful additional information.

We emphasise that these conclusions concerning the positions of the two mangabey lineages should be regarded as suggestive working hypotheses to be tested in terms of the degrees to which they facilitate the interpretation of other bodies of comparative data bearing on cercopithecine evolution. One such test might begin with the observation that the mangabeys have sufficient morphological uniformity to have been placed consistently in a single genus—which is very significant in view of the tendency to splitting evident in Old World monkey taxonomy^{11,23}. This morphological uniformity contrasts sharply with the molecular picture just presented. Such a contrast is no longer surprising in the context of the extensive recent documentation of the independence of rates of structural protein and organismal evolution^{24–27}, but its presence provides the opportunity to develop important insights into the evolution of the Cercopithecinae. If the contrast is real, then either mangabey morphology is primitive for the Papionini (baboons, macaques and mangabeys), or it is derived, in parallel or convergently, from the common ancestor of the group. This latter possibility is rendered unlikely by the fact that no one has yet seen mangabey morphology as derived^{19,20,28,29}. Thus we are left with the former, leading to the important logical corollary that the larger derived Papionini—drill and mandrill, savanna baboons and gelada, and some large macaques such as *M. nigra* and *M. nemestrina*—simply represent slightly different ways of being a large Old World monkey starting from a mangabey-like morphological base.

One of the *albigena* samples (19354) was obtained from the Delta Regional Primate Center (DRPC) through Dr Charles Hill, and the other from eastern Zaire through Dr Urs Rahm. The two were immunologically indistinguishable. *C. aterrimus*

was provided by Dr Kurt Benirschke of the San Diego Zoo; *Papio* and *Theropithecus* by Dr J. Moor-Jankowski of LEMSIP; *Mandrillus*, *C. ascanius* and *C. guereza* by Dr William Mottram of San Francisco Zoo; *C. galeritus* (061-6979, 6984, 8512, 9093), *C. torquatus* (062-6992, 7004, 7191), *C. aethiops* and *C. patas* by DRPC; *C. torquatus* (CFU-2, 14, 26, 28) and *P. entellus* by Davis Regional Primate Center and *Macaca* by Dr Marjorie LaSalle of the Oregon Regional Primate Center. All work was done in the laboratory of Dr Allan C. Wilson of the University of California, Berkeley. We thank him for facilities and helpful discussion, together with S. L. Washburn and W. Malmi. Technical assistance was provided by Anne Hill and Peter Hornbeck and support through grants from the US National Science Foundation to A.C.W. and V.M.S. and a predoctoral fellowship to J.E.C.

JOHN E. CRONIN
VINCENT M. SARICH

Departments of Anthropology and Biochemistry,
University of California,
Berkeley, California 94720

Received November 26, 1975; accepted February 17, 1976.

- ¹ Goodman, M., *Ann. N.Y. Acad. Sci.*, **102**, 219 (1962).
- ² Sarich, V. M., in *Old World Monkeys* (edit. by Napier, J. R., and Napier, P. H.), 175 (Academic, New York, 1970).
- ³ Cronin, J. E., and Sarich, V. M., *J. hum. Evol.*, **4**, 357 (1975).
- ⁴ Sarich, V. M., *Syst. Zool.*, **18**, 286 (1969).
- ⁵ Leone, C. A., and Wiens, A. L., *J. Mammal.*, **37**, 11 (1956).
- ⁶ Sarich, V. M., *Nature*, **245**, 218 (1973).
- ⁷ Maxson, L. R., and Wilson, A. C., *Science*, **185**, 66 (1974).
- ⁸ Goodman, M., in *Classification and Human Evolution* (edit. by Washburn, S. L.) (Aldin., Chicago, 1963).
- ⁹ Goodman, M., and Moore, G. W., *Syst. Zool.*, **20**, 19 (1969).
- ¹⁰ Hoyer, B. H., van deVelde, N. W., Goodman, M., and Roberts, R. E., *J. hum. Evol.*, **1**, 645 (1972).
- ¹¹ Napier, J. R., and Napier, P. H., *A Handbook of Living Primates* (Academic, New York and London, 1967).
- ¹² Gray, A. P., *Mammalian Hybrids*, second ed. (Commonwealth Agricultural Bureaus, Farnham Royal, UK, 1972).
- ¹³ Delson, E., in *Approaches to Primate Paleobiology* (edit. by Szalay, F.), 167 (Karger, Basel, 1975).
- ¹⁴ Cronin, J. E., thesis, Univ. California, Berkeley (1975).
- ¹⁵ Champion, A. B., Prager, E. M., Wachter, D., and Wilson, A. C., in *Biochemical and Immunological Taxonomy of Animals* (edit. by Wright, C. A.), 397 (Academic, London, 1974).
- ¹⁶ Prager, E. M., and Wilson, A. C., *J. biol. Chem.*, **246**, 7010 (1971).
- ¹⁷ Wallace, D. G., King, M. C., and Wilson, A. C., *Syst. Zool.*, **22**, 1 (1973).
- ¹⁸ Maxson, L. R., and Wilson, A. C., *Syst. Zool.*, **23**, 1 (1975).
- ¹⁹ Jolly, C. J., in *The Baboon in Medical Research*, 2 (edit. by Vagtberg, H.), 323 (University of Texas Press, Austin, 1966).
- ²⁰ Jolly, C. J., in *Old World Monkeys* (edit. by Napier, J. R., and Napier, P. H.), 139 (Academic, New York, 1970).
- ²¹ Barnicot, N. A., and Hewett-Emmett, D., *Folia primat.*, **17**, 442 (1972).
- ²² Barnicot, N. A., and Wade, P. T., in *Old World Monkeys* (edit. by Napier, J. R., and Napier, P. H.), 227 (Academic, New York, 1970).
- ²³ Osman Hill, W. C. O., *Primates: Comparative Anatomy and Taxonomy*, 7 (Edinburgh University Press, Edinburgh, 1974).
- ²⁴ Wilson, A. C., Maxson, L. R., and Sarich, V. M., *Proc. natn. Acad. Sci. U.S.A.*, **71**, 2843 (1974).
- ²⁵ Wilson, A. C., Sarich, V. M., and Maxson, L. R., *Proc. natn. Acad. Sci. U.S.A.*, **71**, 3028 (1974).
- ²⁶ Prager, E. M., and Wilson, A. C., *Proc. natn. Acad. Sci. U.S.A.*, **72**, 200 (1975).
- ²⁷ King, M. C., and Wilson, A. C., *Science*, **188**, 107 (1975).
- ²⁸ Simons, E. L., *Primate Evolution* (Macmillan, New York, 1972).
- ²⁹ von Koenigswald, G. H. R., in *Fossil Vertebrates of Africa* (edit. by Leakey, L. S. B.), 39 (Academic, New York, 1969).

Magnesium emitted by snails alters swimming behaviour of *Schistosoma mansoni* miracidia

Schistosoma mansoni is a parasitic flatworm transmitted by snails which infects millions of people in tropical countries with poor sanitation, where new irrigation schemes, population growth and continued poverty have increased the prevalence of infection. Larval worms, emerging from freshwater snails, penetrate human skin and mature in intestinal veins where they lay eggs. Many eggs remain in and damage tissues while others pass out in excreta and hatch in water, releasing embryos called miracidia. These short lived, multicellular, ciliated, free-swimming miracidia must find, penetrate and reproduce within certain snails whence emerge the larvae infective for man. Snails emit substances ("miraxones") which alter the swimming be-

haviour of miracidia and may help them to locate and attack the snails¹⁻³. Miracidial behaviour is reportedly affected by snail-elaborated amino acids^{4,5} and by various other agents, including fatty acids^{6,7}. We now report that magnesium appears in water in which uninfected *Biomphalaria glabrata* (a snail host of *Schistosoma mansoni*) is confined, and that magnesium ions alter miracidial swimming patterns.

Methods for rearing Puerto Rican (PR-2') *B. glabrata* and for obtaining miracidia from livers of mice infected with *S. mansoni* have been described before, as has a semi-quantitative bioassay for miraxone¹. In this assay, miracidial responses are observed stereomicroscopically and scored 1 to 4+ after wells (1.5 cm diameter, each containing 50-100 miracidia in 0.3 ml of commercial spring water) are "point-inoculated" at the edge with 20 μ l of test fluid; 4+ denotes strong, persistent miracidial aggregation and excitation around the zone of inoculation, whereas lower scores denote progressively less pronounced aggregations, as previously detailed¹. Replicate experiments yielded similar results.

We have used the bioassay to detect miraxonal activity at each step of a procedure designed to isolate and purify miraxone(s) from "snail-conditioned water" (SCW) prepared by confining 100 snails (12-16 mm diameter) in 100 ml of food-free distilled water for 5 h at 24 $\pm$ 1 $^{\circ}$ C. SCW was filtered through paper and successive batches were pooled and stored at -10 $^{\circ}$ C until processed. Lyophilising SCW yielded "snail-water powder" (SWP) which was extracted twice with chloroform-methanol, 2:1 (125 ml per g SWP). The pooled extract was shaken with distilled water⁸, and the resulting lipid-free upper phase—containing most of the miraxone originally present in SWP—was separated, dried, dissolved in distilled water, and then passed through a Diaflo UM-05 ultrafiltration membrane (Amicon) retaining material of molecular weight >500. The filtrate was fractionated in Sephadex G-10 gel using distilled water as eluant, and a single band of miraxone was detected at V_e/V_0 (elution volume/void volume)=1.3-2.0. This miraxone was further purified by ascending paper chromatography in two successive solvent systems: (1) butanol-methanol-acetate-water, 4:8:1:1 (R_f , or migration relative to solvent front, of miraxone=0.5-0.8); and (2) isopropanol-pyridine-water, 7:7:2 (R_f of miraxone=0.3-0.55). A sample of 51 of SCW yielded about 1 g SWP from which we obtained about 26 mg miraxone after the full purification process.

When eluted from the second paper chromatogram, the miraxone was soluble in water and alcohol; insoluble in ether, hexane and chloroform; heat stable (120 $^{\circ}$ C for 4 h); stable at all pH levels; strongly deliquescent; odourless, and non-volatile. It formed a glassy film when dried. The product apparently lacked amino groups as it proved negative by ninhydrin spot tests and by fluorescamine analysis⁹. Tests for reducing sugars and fatty acids were negative. Mass spectroscopy to determine molecular weight failed because the material was non-volatile.

The purified miraxone contained 8.9-21.6% magnesium, 28.3-35.4% chlorine, 0-8.5% carbon, 5.8-6.0% hydrogen and 0.5-0.6% nitrogen (% by weight; range of two separately purified samples). Control experiments revealed that (1) distilled water, processed through the purification scheme described above, yielded no miracidial stimulant and only 0.03% of the magnesium present in SCW-purified miraxone, and (2) the carbon and nitrogen in miraxone probably represented contaminants from chromatography.

A grey residue, remaining after purified miraxone was ashed, contained 28% magnesium by weight but only traces of other metals. When this ash was dissolved in 0.5 N HCl, dried, and dissolved in water, its miraxonal activity (as measured at serial twofold dilutions¹) resembled that of the miraxone before ashing. By contrast, when SWP was ashed and dissolved as described, its activity decreased, perhaps

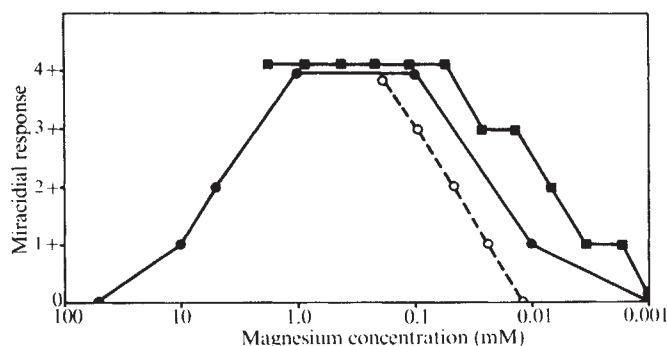


Fig. 1 Responses (1 to 4+) of *S. mansoni* miracidia, plotted against magnesium concentrations in serial dilutions of: commercial magnesium chloride (●); a typical sample of *B. glabrata*-conditioned water (○), and miraxone purified from *B. glabrata*-conditioned water as described (■). Dilutions were all made with distilled water. Magnesium concentrations were determined by atomic absorption spectrophotometry. The decreased miracidial response at concentrations of commercial magnesium chloride above 1 mM represents an unexplained effect also observed with some highly concentrated preparations of miraxone. Tests with magnesium sulphate yielded curves resembling that shown for magnesium chloride.

because ashing may alter the ratio of magnesium to interfering cations (see below).

The purified miraxone migrated 9–12 cm cathodally in high voltage paper electrophoresis (30 min at 3 kV; buffer, pyridine–acetate–water, 400:16:3584; pH 6.42; Whatman 3MM paper); therefore, the miraxone is a cation at this pH. Electrophoresis removed chloride from the miraxone but left the biological activity unchanged.

The above findings suggested that the material purified from SCW is magnesium chloride, and that the biologically active miraxone is magnesium ion. This hypothesis is strengthened by three other lines of evidence. First, aqueous solutions of $MgCl_2$ or $MgSO_4$ (1.0–0.1 mM) evoked strong miracidial responses (Fig. 1), while solutions of $CaCl_2$, KCl , K_2SO_4 and $NaCl$ (1.0–0.001 mM) elicited no miracidial response. Second, when we processed magnesium chloride ($MgCl_2 \cdot 6H_2O$, ACS grade, Fisher Scientific) through the purification scheme (starting with chloroform–methanol extraction), miraxonal activity purified exactly as the SCW-derived miraxone at each step. Finally, $MgCl_2 \cdot 6H_2O$ contains, by calculated weight, 12% magnesium, 35% chlorine and 5.9% hydrogen—values close to those of the purified miraxone.

Several additional findings favour magnesium ion as the major miraxone in SCW. Fresh SCW, which induces 4+ miracidial responses, contained an average of 0.24 mM magnesium (range: 0.17–0.35 mM). More important, the dilution–response curves for SCW and for purified miraxone parallel that of magnesium chloride at comparable magnesium concentrations (Fig. 1). The magnesium content of SCW, therefore, can account for the miraxonal activity of this solution assuming that all or most of the magnesium is chemically ‘available’ as a miracidial stimulant and that other components of SCW do not interfere with miracidial stimulation by magnesium.

To test for other miraxones in SCW, the magnesium concentration in test wells containing miracidia was made up to 0.33 mM by adding $MgCl_2 \cdot 6H_2O$ 5 min before challenging the miracidia with SCW of known activity. Although the miracidia swam normally, they did not respond to the SCW. This result is consistent with magnesium being the only major miraxone in SCW. The increased background magnesium (now greater than that of SCW) would have eliminated any magnesium gradient from the inoculated SCW, but this should not have affected gradients of other miraxones if any such were present. We cannot,

however, deny that magnesium could nonspecifically inhibit miracidial responses to other hypothetical miraxones.

The miraxonal activity of a solution may depend in part on the concentrations of inorganic ions other than magnesium. For example, the activity of SCW¹ and of a 0.47 mM magnesium chloride solution is reduced if 34 mM NaCl is added to these solutions before bioassay. Although we do not understand the basis for this effect, the blockage suggests that direct comparison of the miraxonal activity of purified and crude materials may be misleading because interfering ions or other substances may be present in the crude preparations.

Although we do not exclude the possibility that organic molecules in crude SCW contribute to some of its miraxonal activity, our findings show that magnesium ion accounts for most of that activity. Since previous evidence indicates some nonspecificity between the snail species releasing miraxone(s) and the species of miracidium responding^{1–3}, perhaps the miracidia of many species of trematode respond to magnesium. This response may be a requisite part of the miracidial host-finding process. It remains to be shown how snails release magnesium, how miracidial behaviour is altered by magnesium and how this information can be used to interrupt the transmission of schistosomiasis.

We acknowledge preliminary research by Dr Robert Murphy, and the technical assistance of Ms Christine A. Burns and Ms Nora S. Anderson. This study was supported in part by a research grant, and by a research career award to E.C., from the US Public Health Service.

HENRY H. STIBBS

ELI CHERNIN

Department of Tropical Public Health,
Harvard School of Public Health,

SAMUEL WARD

MANFRED L. KARNOVSKY

Department of Biological Chemistry,
Harvard Medical School,
Boston, Massachusetts 02115

Received December 31, 1975; accepted February 24, 1976.

- Chernin, E., *J. Parasit.*, **56**, 287–296 (1970); *ibid.*, **58**, 209–212 (1972); *Am. J. trop. Med. Hyg.*, **23**, 320–327 (1974).
- Ulmer, M. J., in *Ecology and Physiology of Parasites* (edit. by Fallis, A. M.), 123–160 (University of Toronto Press, Toronto, 1971).
- Cable, R. M., in *Behavioural Aspects of Parasite Transmission* (edit. by Canning, E. U., and Wright, C. A.), *Zool. J. Linnean Soc.*, **56**, Suppl. No. 1, 1–18 (Academic, London, New York, 1972).
- Wright, D. G. S., and Ronald, K., *Can. J. Zool.*, **50**, 855–860 (1972).
- MacInnis, A. J., Bethel, W. M., and Cornford, E. M., *Nature*, **248**, 361–363 (1974).
- MacInnis, A. J., *J. Parasit.*, **51**, 731–746 (1965).
- Wilson, R. A., and Denison, J., *Comp. Biochem. Physiol.*, **32**, 511–517 (1970).
- Folch, J., Lees, M., and Sloane-Stanley, G. H., *J. biol. Chem.*, **226**, 497–509 (1957).
- Weigle, M., DeBernardo, S., Teng, J., and Leimgruber, W., *J. Am. chem. Soc.*, **94**, 5927–5928 (1972).

Temperature dependence of biochemical oscillations in cell suspensions of *Dictyostelium discoideum*

ON a substratum the aggregation of amoebae of *Dictyostelium discoideum* often occurs in pulses because the aggregation centres release the chemotactic signal periodically^{1,2}. The signal is believed to be cyclic AMP^{3,4}. Cyclic AMP oscillations have been observed in suspensions of aggregation competent cells of *D. discoideum*⁵. Since circadian rhythms are largely compensated for temperature effects⁶ the temperature dependence of other biological rhythms is of general interest.

In suspensions of *D. discoideum* cells, changes in absorbance accompanying periodic cyclic AMP production can be recorded^{5,7}. These changes occur as an initial series of spikes followed by sinusoidal oscillations⁷. We have studied the sinusoidal oscillations since they were found to

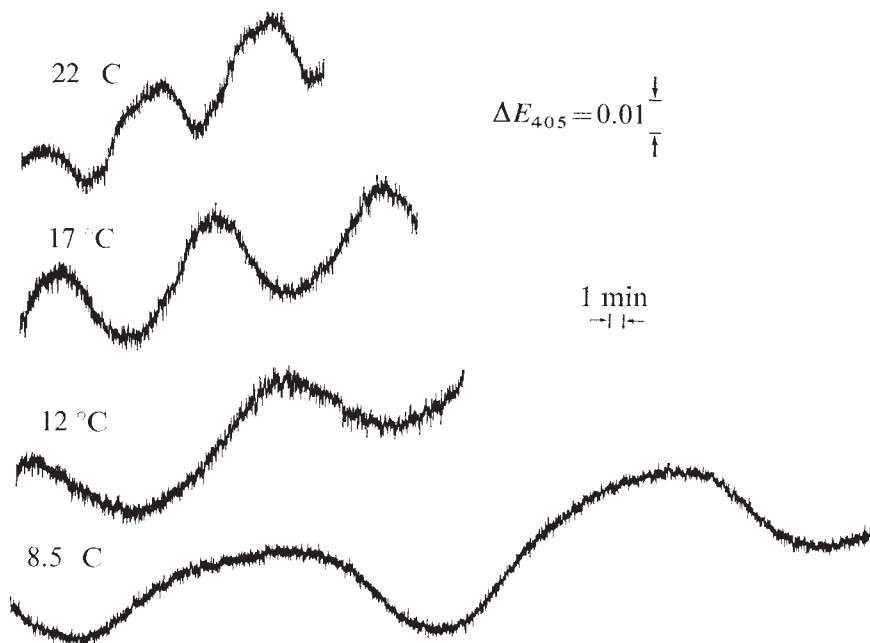


Fig. 1 Oscillations in cell suspensions of *D. discoideum* at different temperatures. Amoebae of the axenic strain Ax-2 (clone 214) were cultivated at 23 °C on nutrient medium supplemented with 1.8% maltose¹¹ and collected at cell densities of 0.2 to 1×10^7 ml⁻¹. The cells were washed three times in the cold with 0.017 M Sørensen phosphate buffer, pH 6.0, adjusted to 1×10^7 cells ml⁻¹ in buffer and shaken for 4 h at 23 °C. After this the amoebae were centrifuged, adjusted to 5×10^7 cells ml⁻¹ in buffer, transferred into a cuvette with an optical path of 1 cm, and agitated by bubbling oxygen at a flow rate of 24 ml min⁻¹ through the suspension. With a Zeiss PM6 spectrophotometer the optical density at 405 nm was recorded. Changes in absorbance seem to be due to changes in light scattering⁷ and indicate changes in shape or size of cells. The absorbance decreases towards the top. The starting temperature was 22 °C and after sinusoidal oscillations began (after ~1–2 h) the temperature was either kept at 22 °C or slowly shifted to 17 °C, 12 °C, or 8.5 °C. In all experiments the actual temperature in the cuvette was determined.

proceed for more periods than the spikes. Sinusoidal oscillations of amoebae of *D. discoideum* strain Ax-2 (clone 214) have been observed in the temperature range from 7 °C to 24 °C. As shown in Fig. 1, by decreasing the temperature from 22 °C to 8.5 °C the oscillation period increases from 7.8 to 27 min. The dependence of the oscillation period τ on the temperature T can be described by the Arrhenius equation (Fig. 2), and an activation energy for the oscillating system of $E=16$ kcalorie is obtained.

Temperature does not affect the form of sinusoidal oscillations (Fig. 1). At any temperature the part of the wave from the maximum to the minimum in absorbance is longer than that from the minimum to the maximum. The amplitude does not decrease with decreasing temperature;

on the contrary, in some experiments higher amplitudes were obtained at lower temperature.

During the course of these studies we learned from Nanjundiah that he and his colleagues⁸ have investigated the temperature dependence of the frequency of periodic movement steps during the aggregation of amoebae of *D. discoideum* strain NC-4H on agar. Although at any temperature the oscillation period of NC-4H amoebae on agar is shorter than that of Ax-2 amoebae in suspension, the activation energy is the same under both conditions: 15.8 kcalorie (ref. 8) compared with 16 kcalorie. This result is in accord with the assumption that the same biochemical oscillator underlies rhythmic signalling of aggregation centres on agar as well as periodic activities in cell suspensions.

Nanjundiah *et al.*⁸ interpreted E as the activation energy of a rate-limiting step in the biochemical oscillation. Alternatively a linear Arrhenius plot for a network of reactions results if the activation energies of all individual steps are similar. The activation energy of 16 kcalorie is in the range of those calculated for enzyme-catalysed reactions⁹. A non-enzymatic chemical oscillator, the Belousov–Zhabotinskii system, shows an even stronger temperature dependence of the oscillation period¹⁰, corresponding to an activation energy of 26 kcalorie (as calculated from the data of ref. 10).

I thank Drs G. Gerisch, F. Grün and V. Nanjundiah for stimulating discussions. This work was supported by the Schweizerische Nationalfonds.

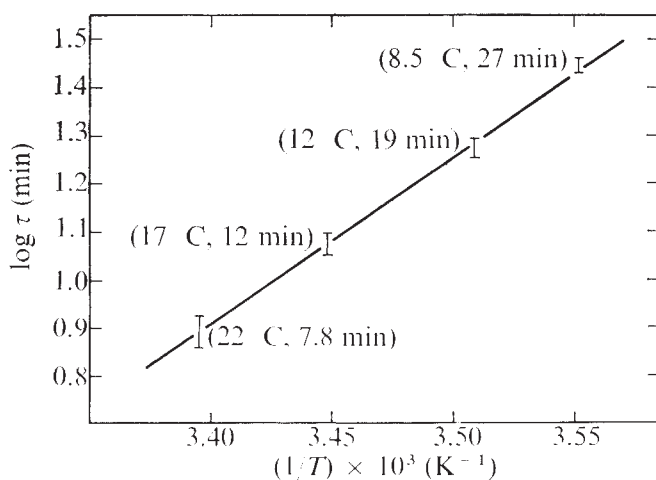
BERND WURSTER

Biozentrum der Universität,
4056 Basel, Switzerland

Received February 4; accepted February 20, 1976.

- ¹ Shaffer, B. M., *Adv. Morphog.*, **2**, 109–182 (1962).
- ² Gerisch, G., *Curr. Top. dev. Biol.*, **3**, 157–197 (1968).
- ³ Konijn, T. M., van de Meene, J. G. C., Bonner, J. I., and Barkley, D. S., *Proc. natn. Acad. Sci. U.S.A.*, **58**, 1152–1154 (1967).
- ⁴ Bonner, J. T., Barkley, D. S., Hall, E. M., Konijn, T. M., Mason, J. W., O'Keefe, G., and Wolfe, P. B., *Dev. Biol.*, **20**, 72–87 (1969).
- ⁵ Gerisch, G., and Wick, U., *Biochem. biophys. Res. Commun.*, **65**, 364–370 (1975).
- ⁶ Bünning, E., *Die Physiologische Uhr* (second ed.) (Springer, Berlin, Göttingen, Heidelberg, 1963).
- ⁷ Gerisch, G., and Hess, B., *Proc. natn. Acad. Sci. U.S.A.*, **71**, 2118–2122 (1974).
- ⁸ Nanjundiah, V., Hara, K., and Konijn, T. M., *Nature*, **260**, 705 (1976).
- ⁹ Dixon, M., and Webb, E. C., *Enzymes* (second ed.) (Longmans, London, 1964).
- ¹⁰ Lefelhoc, J. F., *J. chem. Educ.*, **49**, 312–314 (1972).
- ¹¹ Watts, D. J., and Ashworth, J. M., *Biochem. J.*, **119**, 171–174 (1970).

Fig. 2 Dependence of the oscillation period on the temperature. The Arrhenius equation was applied in the form $\log \tau = (E/2.3RT) + \text{constant}$ (where τ is the oscillation period, R the gas constant, and T the absolute temperature) and $\log \tau$ was plotted against T^{-1} . The Celsius temperature and the period of oscillation are given in parentheses. At 22 °C sinusoidal oscillations of Ax-2 amoebae in suspension ceased after about 20 periods, and the period length decreased from the beginning to the end by 10 to 20%. At each temperature the period length was determined between 5 and 10 oscillation periods from the beginning. The data represent the range of period lengths evaluated from at least 4 oscillation periods at each temperature.



Effect of temperature on morphogenetic oscillations in *Dictyostelium discoideum*

WHEN their food supply is exhausted, single amoebae of the cellular slime mould *Dictyostelium discoideum* aggregate¹. On a two-dimensional substrate such as agar, this aggregation is the result of a succession of periodic coordinated movement steps directed towards a group of amoebae—the “centre”, which may in fact lie anywhere in the field of cells. Recent theoretical^{2,3} and experimental^{4–7} work suggests that this process can be explained as follows. (1) The centre produces and releases cyclic AMP (the signal) in an oscillatory manner, and (2) other cells in the field respond chemotactically by moving up cyclic AMP gradients and relay the signal progressively outwards. By means of time-lapse cinematography we have looked at the oscillation frequency ν during the aggregation of small drops (about 1,000 cells in a drop of diameter 1 mm) of *D. discoideum* cells (strain NC-4H) aggregating on hydrophobic agar⁸ at various temperatures, and have found that the oscillation frequency increases with temperature (Fig. 1).

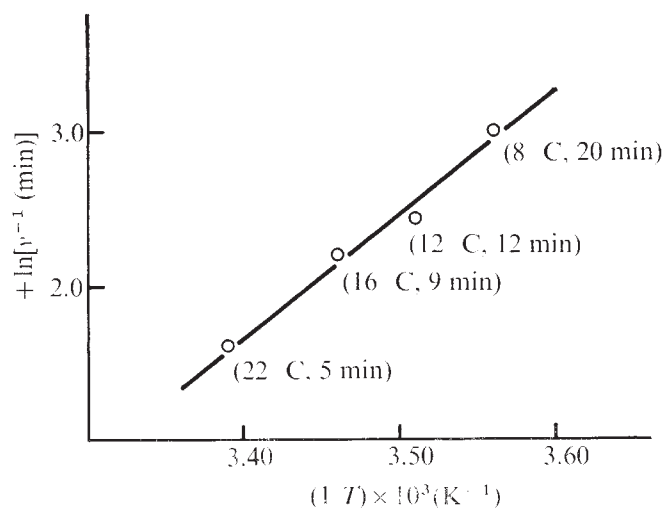


Fig. 1 T is the absolute temperature and ν the frequency of oscillation. Experimental points are fitted to the curve $\ln(\nu^{-1}) = \ln(\nu_0^{-1}) + (\Delta E/RT)^{-1}$, with R equal to the gas constant. Figures in parentheses indicate the temperature (°C) and period of oscillation. The amoebae were deposited in a drop of Bonner's salt solution¹¹ on agar, and the agar plate was kept in the temperature-controlled room for 2–3 h before aggregation began. Temperature varied within $\pm 0.5^\circ\text{C}$ of the recorded means. Periods were recorded as differences in times of initiation of wave propagation at the centre, each such initiation having the appearance of a quick contraction. Mean frequencies were estimated from two experiments at each temperature with about 15 oscillations recorded during each aggregation. A sample run of periods, recorded at 12°C , reads 13, 12, 12, 13, 13, 12, 12, 12, 11, 12 min. The aggregates were concentric, that is, morphologically similar in all experiments.

These amoebae aggregate and form viable spores (as tested at 22°C) down to 8°C . The Arrhenius plot ($\ln \nu^{-1}$ against $\ln T^{-1}$) is linear, corresponding to a frequency factor $\nu_0 = 2 \times 10^9 \text{ s}^{-1}$ and an activation energy $\Delta E = 15.8 \text{ kcal mol}^{-1}$. ΔE can be interpreted as the activation energy of a (hypothetical) rate-limiting step in the overall scheme for the biochemical oscillation. It is noteworthy that this value of ΔE is in the range of activation energies for non-transport enzyme-catalysed reactions⁹, whereas those for known specific transport processes are much smaller¹⁰. In parallel with the increase in oscillation frequency, there is also an increase in the mean speed of amoeboid move-

ment (not shown). Thus, other factors being the same, the number of pulses needed for a field of cells to aggregate ought to be only weakly temperature dependent. In fact, if we consider just those cases (four in all) where oscillations could still be observed till late in aggregation, this number was 20 at 12°C , 19 at 16°C and 17 at 22°C . Rigorous substantiation of this last result would suggest that there is an underlying oscillator controlling movement, possibly coupled to the one involved in signal generation.

VIDYANAND NANJUNDIAH

Biozentrum der Universität,
4056 Basel, Switzerland

KOKI HARA

Hubrecht Laboratory,
Uppsalalaan 8,
Utrecht

THEO. M. KONIJN

Zoölogisch Laboratorium,
Rijksuniversiteit te Leiden,
The Netherlands

Received December 4, 1975; accepted February 20, 1976.

- 1 Bonner, J. T., *The Cellular Slime Molds* (Princeton University Press, New Jersey, 1967).
- 2 Goldbeter, A., *Nature*, **253** (5492), 540–542 (1975).
- 3 Nanjundiah, V., *J. theor. Biol.*, **56**, 275–282 (1976).
- 4 Gerisch, G., and Wick, U., *Biochem. biophys. Res. Commun.*, **65**(1), 364–370, (1975).
- 5 Roos, W., Nanjundiah, V., Malchow, D., and Gerisch, G., *FEBS Lett.*, **53**(2), 139–142 (1975).
- 6 Shaffer, B. M., *Nature*, **255**, 549–552 (1975).
- 7 Mato, J. M., Losada, A., Nanjundiah, V., and Konijn, Th. M., *Proc. natn. Acad. Sci. U.S.A.*, **72**, 4991–4993 (1975).
- 8 Konijn, Th. M., and Raper, K. B., *Dev. Biol.*, **3**, 725–756 (1961).
- 9 Dawes, E. A., in *Comprehensive Biochemistry*, **12** (edit. by Florkin, M., and Stotz, E. H.) (Elsevier, New York, 1964).
- 10 Stein, W. D., *The Movement of Molecules Across Cell Membranes* (Academic, New York, 1967).
- 11 Bonner, J. T., *J. exp. Zool.*, **106**, 1–26 (1947).

External cyclic AMP-dependent protein kinase activity in rat C-6 glioma cells

IN certain eukaryotic cells β -agonistic catecholamines bind to membrane-associated hormone receptors, and stimulate the intracellular synthesis of cyclic AMP by activation of the adenylate cyclase¹. In C-6 rat glioma cells stimulated with noradrenaline this increase of cyclic AMP leads to, or is correlated with, a morphological change in the cells². When the intracellular cyclic AMP concentration has reached a maximum, it returns to basal values³ due to an energy-dependent excretion of cyclic AMP from the cell into the external medium^{3,4}, similar to that found in several mammalian systems^{5–7}. It is possible that the excreted cyclic AMP interacts either with an external, glial-membrane bound cyclic AMP receptor or with the surface of the surrounding cells, for example neurones. We report here that intact, cultured glial cells seem to have a cyclic AMP-dependent protein kinase on their external surface which transfers γ -phosphate of ATP to an external acceptor protein (histone). This protein kinase activity is stimulated indirectly by the addition of noradrenaline to the cells. The physiological significance of this process is unknown.

We first tested the effect of extracellular cyclic AMP on the ability of a protein kinase activity of intact C-6 cells to phosphorylate an exogenous substrate. Figure 1 shows the phosphorylation by C-6 cells of added external histones as a response to incubation in the presence of $5 \times 10^{-6} \text{ M}$ cyclic AMP. These data suggest that actively growing glial cells have a cyclic AMP-dependent protein kinase, probably on the external cell surface, which can be stimulated from the outside of the membrane. The same experiment showed that a small fraction of ^{32}P radioactivity remained with the cells and that there was no significant difference between stimulated and unstimulated cells with respect to this accumulation (data not shown).

There are two possible objections to the experiment

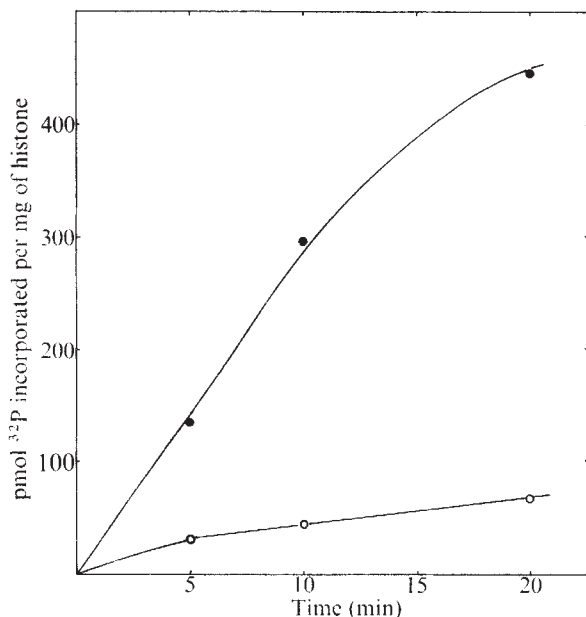


Fig. 1 Activation of external protein kinase-mediated incorporation of ^{32}P into histone. C-6 rat glioma cells (flow laboratories), originally derived by Benda *et al.*⁹, were grown routinely in 60-mm Petri dishes with Ham's F-10 medium supplemented with 15% horse serum, 2.5% foetal bovine serum and gentamicin ($50 \mu\text{g ml}^{-1}$) in an atmosphere of 90% air and 10% CO_2 at 37°C . All experiments were performed with confluent cultures (about $3\text{--}4 \times 10^6$ cells per plate) since it has been shown that cell-cell contact is required for maximal increase of internal cyclic AMP after hormone treatment¹⁰. To measure the external protein kinase activity, the cells were first washed three times with prewarmed Dulbecco's phosphate-buffered saline (PBS), then 1 ml of reaction buffer (50 mM MOPS buffer, pH 7.2, 5 mM MgCl_2 , 0.125 mM CaCl_2 , 125 mM sucrose, 7 mM mercaptoethanol) was added to start the incubation. The reaction buffer included $\gamma\text{-}^{32}\text{P}\text{-ATP}$ (specific activity 1.3×10^3 c.p.m. pmol^{-1}), prepared according to Glynn and Chappel¹¹, and calf thymus histones ($300 \mu\text{g ml}^{-1}$, Calbiochem). If the effect of cyclic AMP was to be measured, the reaction buffer also contained 5×10^{-6} M cyclic AMP. During the experiment the dishes were placed on a constantly moving, shaking table. After the periods indicated at 35°C , the buffer was removed and the cells were washed with 1 ml of cold PBS. Buffer and washings were combined and centrifuged (2 min, $1,500g$), and protein in the supernatant was precipitated with 10% trichloroacetic acid (TCA). The precipitates were collected on Whatman GF/A glass fibre filters. The filters were washed six times with 5 ml of 5% TCA and once with 3 ml of methanol, dried and counted in a liquid scintillation counter. The cells were washed three times with cold PBS and lysed with 1 ml of 0.1% sodium dodecyl sulphate. The cell lysates (1 ml) were counted in 10 ml of a toluol-Triton X-100 mixture (9 parts toluol+1 part Triton X-100 containing 0.5% PPO). Each point represents the mean of three determinations: ●, 5×10^{-6} M cyclic AMP; ○, without cyclic AMP.

shown in Fig. 1. Perhaps the cyclic AMP added to the culture medium penetrates into the cells and activates an internal protein kinase. This is unlikely because it seems improbable that cyclic AMP enters glioma cells—it does not induce the characteristic morphological alterations which are a prompt response of these cells to the addition of dibutyryl cyclic AMP or catecholamines (our unpublished observations). It is also possible that a small fraction of the cells disintegrate during incubation. Cytoplasmic cyclic AMP-dependent protein kinase could then phosphorylate the added histones. This possibility is also unlikely because microscopy has revealed no gross alteration of cellular integrity. Contamination is ruled out by the experiments summarised in Table 1. These controls show that less than 2% cyclic AMP-dependent protein kinase activity was found in the incubation buffer after contact with the cells. The time course shown in Fig. 1 also excludes the possibility of contamination. The external protein kinase is

Table 1 Control experiments

	pmol ^{32}P incorporated per mg of histone per 10 min	
	Reaction buffer without cyclic AMP	Reaction buffer with 5×10^{-6} M cyclic AMP
Cell-free supernatant	5.05 ± 0.85	5.25 ± 0.44
Standard reaction procedure (with cells)	38.42 ± 2.24	302.56 ± 4.37

Each value represents the mean $\pm$ s.e. of five determinations.

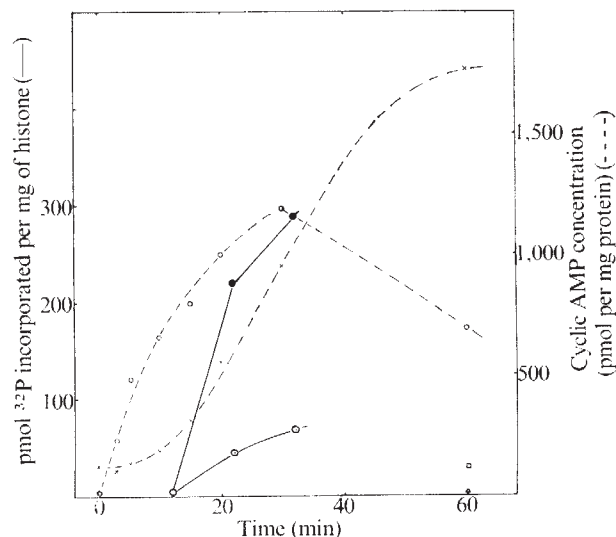
To show that the external protein kinase activity is really located at the cell surface we added the ^{32}P acceptor protein to the reaction buffer after having incubated the buffer with growing cells. Incubation buffer without histones was removed from the cells after 10 min of incubation at 35°C in the conditions described for Fig. 1. The sample was centrifuged for 5 min at $1,500g$ and protein kinase activity in the cell-free supernatant was determined after addition of histones. The radioactivity in TCA-precipitated histones was determined after 10 min of incubation as described for Fig. 1 (cell-free supernatant). These values are compared with values obtained when histones were present with the cells during the first 10-min incubation and the second incubation not performed (standard reaction procedure).

most active at cyclic AMP concentrations higher than 10^{-6} M and has a pH optimum of about 7.2 (data not shown).

Secretion of cyclic AMP from C-6 rat glioma cells into the medium was originally described as a response to treatment with L-isoproterenol^{3,4}. We show here that noradrenaline also causes secretion of cyclic AMP from glioma cells.

Figure 2 shows the intracellular and extracellular levels

Fig. 2 Activation of external protein kinase by noradrenaline. Noradrenaline (10^{-4} M, Sigma) was added to confluent C-6 glioma cells growing in 60-mm dishes with 1 ml of medium. At the times indicated, cyclic AMP was extracted with 5% TCA as described by Oey² and cyclic AMP concentration was determined according to Gilman¹². For protein determination¹², bovine serum albumin (Sigma) was used as standard. To measure the ability to transfer ^{32}P into histone in the presence of noradrenaline the cells were preincubated in Ham's F-10 medium for 12 min. Then the medium was replaced with incubation buffer containing 10^{-4} M noradrenaline and the protein kinase activity was measured as described for Fig. 1. Each value represents the mean of duplicate determinations on two cell populations. —●—, Protein kinase activity in the presence of noradrenaline; —○—, in the absence of noradrenaline; --○--, internal cyclic AMP concentration in the presence of noradrenaline; ×, external cyclic AMP concentration with noradrenaline; △, internal cyclic AMP concentration without noradrenaline; □, external cyclic AMP concentration without noradrenaline. The cyclic AMP efflux from growing C-6 cells into incubation buffer occurs at nearly the same rate as into medium (data not shown).



of cyclic AMP after incubation in the presence of 10^{-4} M noradrenaline. After a short lag period, cyclic AMP appeared in the external medium. The external protein kinase could also be activated by treatment of the cells with noradrenaline. This is most clearly demonstrated by the experiments shown in Fig. 2.

The best known and probably the main receptor for cyclic AMP in eukaryotic cells is the regulatory subunit of a protein kinase. After binding cyclic AMP, the regulatory subunit is released from the catalytic subunit, thereby stimulating the activity of the enzyme⁸.

We thank Professor R. Knippers and Drs T. Lindl and A. Powling for helpful discussions. This work was supported by a grant from the Deutsche Forschungsgemeinschaft.

ERNST-JÜRGEN SCHLAEGER
GEORG KÖHLER

Fachbereich Biologie,
Universität Konstanz,
D-7750 Konstanz, West Germany

Received December 31, 1975; accepted March 10, 1976.

- ¹ Robison, G. A., Butcher, R. W., and Sutherland, E. W., *A. Rev. Biochem.*, **37**, 149-174 (1968).
- ² Oey, J., *Nature*, **257**, 317-319 (1975).
- ³ Penit, J., Jard, S., and Benda, P., *FEBS Lett.*, **41**, 156-160 (1974).
- ⁴ Doore, B. J., Bashor, M. M., Spitzer, N., Mave, R. C., and Saier, M. H., *J. Biol. Chem.*, **250**, 4371-4372 (1975).
- ⁵ Cramer, H., and Lindl, T., *Nature*, **249**, 380-382 (1974).
- ⁶ Kuster, J., Zapf, J., and Jakob, A., *FEBS Lett.*, **32**, 73-77 (1973).
- ⁷ Catt, K. J., Watanabe, K., and Dufau, M. L., *Nature*, **239**, 280-281 (1972).
- ⁸ Rubin, C. S., and Rosen, O. M., *A. Rev. Biochem.*, **44**, 831-887 (1975).
- ⁹ Benda, P., Lightbody, J., Sato, G., Levine, L., and Sweets, W. H., *Science*, **161**, 370-371 (1968).
- ¹⁰ Schwarz, J. P., Morris, N. R., and Breckenridge, B. M., *J. Biol. Chem.*, **248**, 2699-2704 (1973).
- ¹¹ Glynn, I. M., and Chappel, J. B., *Biochem. J.*, **90**, 147-149 (1964).
- ¹² Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J., *J. Biol. Chem.*, **193**, 265-275 (1951).
- ¹³ Gilman, A. G., *Proc. natn. Acad. Sci. U.S.A.*, **67**, 305-312 (1970).

H-2 region product as determinant in immune cytotoxicity of syngeneic tumour cells by anti-MSV T lymphocytes

THE existence of cytotoxic T lymphocytes (CTL) specific for antigenically related tumour cells during oncornavirus-induced oncogenesis in the primary host is well documented. They exist in murine sarcoma virus (MSV) tumour-bearing mice, during and after the tumour rejection¹, and in Friend virus-infected mice, notably in the spleen of resistant C57BL/6 (ref. 2). It is probable that they play some role in tumour cell destruction *in vivo*³ and it can be supposed that the antigens of the tumour cell surface recognised by CTL are at least, for the most part, viral proteins or glycoproteins. It was suggested that CTL reacting with virus producer cells in non-oncogenic systems⁴, or with hapten-modified cells^{5,6} recognise an H-2-modified antigen. Therefore, it was of interest to determine whether the same situation exists in an oncornavirus tumour system. It is already known that anti-MSV CTL are much more efficient against syngeneic than against allogeneic tumour cells^{7,8}, indicating that H-2 antigens may play some role in the CTL-target cell interaction. We report here that the H-2 specificities of the tumour cell surface determine the interaction with syngeneic anti-MSV CTL, the reacting antigen being either an H-2 specificity modified by the virus, or, less probably, a virus-induced antigen modified by an H-2 region product.

Anti-MSV CTL derived from different inbred lines and their F₁ hybrids or congenic partners were tested against tumour cells of various H-2 specificities. CTL were taken from the spleen 10-16 d after virus inoculation in BALB/c (H-2^d), B10.D2 (H-2^d), BALB.B (H-2^b), C57BL/6 (H-2^b) and (BALB/c × C57BL/6)F₁ (H-2^{b/d}). All mice were from our colony except BALB.B (Dr F. Lilly, Albert Einstein College, Bronx, New York). BALB/c (C) and BALB.B

Table 1 Specific chromium release (%) by MBL2 (B6) and LSTRA (C) tumour cells attacked by anti-MSV CTL from different strains

Anti-MSV spleen cells from:	Target cells			
	MBL2		LSTRA	
	Expt 1	Expt 2	Expt 1	Expt 2
BALB/c (C)	7	0	17	34
C57BL/6 (B6)	34	26	7	0
(C × B6) F ₁	39	—	16	—
BALB.B	45	—	8	—
B10.D2	7	—	24	—

differ only at the H-2 region of group linkage IX (chromosome 17). A similar situation exists between C57BL/6 (B6) and B10.D2, the latter being congenic to C57BL/10, closely related to B6. Tumour cells were YC8 and LSTRA in C, and MBL2 and FBL3 in B6. YC8, LSTRA and MBL2 were Moloney virus-induced lymphomas; whereas FBL3 was induced by Friend virus. All of them are FMRG(+) cells as defined serologically. MBL2 and FBL3 reacted similarly with B6 anti-MSV CTL as do LSTRA and YC8 with C anti-MSV CTL. The CTL activity was tested using the chromium release test (CRT) carried out at 100:1 CTL-target cells ratio with an incubation period of 6-16 h as described previously¹. It was demonstrated earlier that in the MSV system the cytolytic activity is entirely due to T cells^{1,9}. Results (Table 1) show that B6 anti-MSV CTL are strongly cytolytic for B6 and only weakly for C targets, the reverse situation being observed with C anti-MSV CTL. (B6 × C)F₁ CTL which possess both H-2^d and H-2^b determinants of their parental strains are fully effective against H-2^d and H-2^b targets. CTL obtained from congenic lines behave exactly like their H-2 identical partner: BALB.B CTL lyse B6 better than C tumour cells, whereas B10.D2 are mainly effective against C cells. The use of congenic lines is particularly demonstrative of the H-2 participation in CTL-induced cytotoxicity since they differ from their congenic partners only by the H-2 region. In experiment 1 a weak CTL activity exists for allogeneic targets, also found occasionally in other unpublished experiments. An H-2 incompatibility does not therefore completely abolish the CTL reaction, but as a rule the allogeneic reaction remains very weak as shown in quantitative experiments (F. Plata *et al.*, unpublished). One can therefore conclude that H-2 specificities are determinant in the reaction, H-2 compatibility between CTL and the target being necessary for a fully efficient immune cytotoxicity. Two hypotheses can be proposed to explain this phenomenon: (1) CTL recognise MSV antigens by a specific determinant, but they need at the same time an identical H-2 on their surface and on the target cell to become fully cytotoxic. In other words two different structures are involved separately on the tumour cell: H-2 antigens and the MSV specificities; (2) anti-MSV CTL recognise on the tumour cell surface only one structure which is either an H-2 antigen modified by a viral product, or a viral antigen modified by an H-2 region product. According to the first hypothesis, CTL from F₁ hybrids bearing the H-2 specificities of their two parental lines, F₁ immune CTL must react with any parental cells possessing the MSV-induced antigens. On the contrary, if the second hypothesis were correct, F₁ CTL would react only with target cells sharing an identical H-2 with the immunising parental tumour cells.

(C × B6)F₁ were therefore immunised either by MSV inoculation (which induces F₁ transformed cells *in vivo*) or by 10⁵ MBL2 or FBL3 (B6) or LSTRA (C) tumour cells given subcutaneously. CTL were taken from the spleen about 2 weeks later and tested using the CRT against B6 and C targets. Results show (Table 2) that hybrids immunised against the virus produce CTL reacting with both MBL2 and YC8. On the contrary, hybrids immunised against C or B6 tumour cells produce CTL which react only

Table 2 Cytolytic activity of (C×B6)_F₁ CTL for B6 and C tumour cells (% specific chromium release)

Immunisation by:	Target cells		
	MBL2 (B6)	YC8 (C)	LSTRA (C)
MSV inoculation	38	—	16
	13	8	—
	36	21	—
LSTRA (C) tumour cells	6	—	34
	4	—	26
	0	—	12
MBL2 (B6) tumour cells	7	0	—
FBL3 (B6) tumour cells	13	0	—
	12	0	—

with the immunising cells or antigenically related cells of the same line. One can conclude that H-2 compatibility between CTL and the target is not enough to permit immune cytotoxicity. Hybrids (*H-2^{b/d}*) which received the virus and developed MSV tumour cells have been immunised with MSV antigens associated with (or modified by) H-2^b and H-2^d on the tumour cell surface. Therefore, they react with YC8 (*H-2^d*) as well as MBL2 (*H-2^d*) lymphoma cells. On the other hand, hybrids which received parental tumour cells have been immunised only against tumour antigens associated with the H-2 specificity of the immunising line, and do not therefore react with the tumour antigen of the other parental line. At this stage in the experiments it seems highly probable that anti-MSV CTL recognise only one structure on the target cell surface—an antigen which is different according to the H-2 of the host.

This structure could be an H-2-modified antigen due to the interaction of viral proteins or glycoproteins with H-2 molecules on the tumour cell surface, as proposed for non-oncogenic virus systems⁴ and Hapten-modified cells^{5,6}. According to this hypothesis H-2 and MSV antigens must be closely associated and a preincubation of the targets with related H-2 antibodies ought to block the activity of anti-MSV CTL. Another possibility would be that the virus-induced antigens could be modified by a product of the H-2 region, as proposed in experiments using CTL directed against non-H-2 specificities on the normal cell surface¹⁰. In this case H-2 and MSV antigens are not necessarily associated on the tumour cell surface, and anti-H-2 antibodies would not necessarily block activity. To demonstrate a close topographical relationship between H-2 and MSV antigens, C anti-B6 and B6 anti-C antisera prepared by repeated inoculations of normal spleen cells. These sera were used either unabsorbed or after absorption *in vivo* on congenic spleen cells to provide H-2 and non-H-2 typing sera. For example, C anti-B6 absorbed on BALB.B provides a 'non-H-2 typing serum', whereas the same sera absorbed on B10/D2 provides an 'H-2 typing serum'. Target cells (2×10⁴) were incubated for 30 min at 4 °C in 0.2 ml 1/10 diluted typing serum, then used in the CRT as above. Controls were incubated either with normal mouse serum or with unrelated anti-H-2 typing serum. Results (Table 3) show that unabsorbed and anti-H-2 typing sera block the activity of anti-MSV CTL, whereas non-H-2 typing sera are devoid of blocking activity. Furthermore, a BALB/c anti-BALB.B typing serum which reacts only with H-2 is fully blocking. Anti-Thy1-2 (AKR anti-C3H/He) incubated with the Thy1-2(+) MBL2 lymphoma cells did not inhibit their cytotoxicity by anti-MSV CTL. In addition, unrelated anti-H-2 (for example, B6 anti-C on B6 cells attacked by B6 anti-MSV CTL) are not blocking. It is clear therefore that anti-H-2 have a definite and specific blocking activity for anti-MSV CTL. This result indicates a close topographical relationship between H-2 and MSV antigens and favours the hypothesis of an H-2-modified specificity at the surface of oncornavirus-infected cells. One cannot definitely rule out the hypothesis, however, of a steric hindrance by anti-H-2 antibodies, due to the proximity of H-2 and MSV antigens,

without structural relationship between them. Note also that repeated washings of the target cells after incubation with anti-H-2 clearly altered their blocking activity. It is therefore not possible to draw a definite conclusion from these experiments.

H-2 specificities are clearly determinants in the interaction of anti-MSV CTL and syngeneic tumour cells. Again the H-2 region seems to have a major role in an immunosurveillance phenomenon as reported previously for non-oncogenic virus infections, including lymphocytic choriomeningitis⁴, vaccinia¹¹ and ectromelia¹², after immunisation by hapten-modified cells^{5,6}; and in the immune response against non-H-2 antigens of normal lymphoid cells¹⁰. Similar results have also been found in tumours induced by Friend leukaemia virus¹³ and during *in vitro* stimulation of anti-MSV CTL by tumour cells (F. Plata, personal communication). It is generally assumed that the antigen reacting with CTL is an H-2 modified specificity^{4,6,11,12}. MSV and Friend systems are especially interesting in this case, since the viral polypeptides reacting with H-2 molecules can probably be determined. Further studies are in progress to try to clarify this point and to define the H-2 products involved: H-2D, H-2K or Ia determinants. Preliminary results show that H-2D antigens are especially important.

Another hypothesis must be considered. It has been shown that an H-2 compatibility between CTL and targets is necessary in the CTL reaction directed against non-H-2 antigens of normal lymphoid cells¹⁰. In this situation H-2 and non-H-2 determinants are presumably not on the same structure, and the hypothesis of an H-2 region product specifically modifying non-H-2 molecules must therefore be considered¹⁰. The same could be true in the MSV system: it cannot be excluded, for example, that the antigenic determinants of the major glycoprotein 69/71 of the virus could be modified by an H-2 region-directed molecule. Therefore the nature of the relationship between H-2 and virus antigens is not clearly defined, at least in the MSV system.

Whatever the nature of the H-2/MSV interactions, the existence of an H-2 barrier to the action of anti-virus producer cell CTL merits two comments of practical value. (1) The antigens reacting with CTL in MSV tumour-bearing mice are not completely understood since FMRGi, due to the Moloney virus^{3,8}, and MEV.SA1 antigen, due to endogenous viruses¹⁴, are involved. From the experiments with H-2 congenic lines we can assume that, if an endogenous virus antigen is the target of the anti-MSV CTL it must be integrated in the H-2 region and it has followed this region during the selection of BALB.B and B10.D2 congenic lines. Another, less probable, hypothesis would be that two endogenous viruses of different antigenicity are present both in BALB/c and in B6, their expression being controlled separately by different locus of the H-2 region. It is more likely that FMRGi or other antigens, due to the MSV or to its helper virus, are the main targets of CTL, their specificities being modified according to the H-2 of the host. Note, however, that an H-2 incompatibility strongly decreases but does not completely suppress tumour

Table 3 Blocking of anti-MSV CTL induced cytotoxicity by preincubation of anti-H-2 antibodies with the target cells

Antisera	Specific chromium release by MBL2 targets (%)
C normal serum	18
C anti-B6	0
C anti-B6 absorbed in BALB.B	17
C anti-B6 absorbed in B10.D2	4
C anti-BALB.B	0
B6 anti-C	19
AKR anti-Thy 1-2	19

cells lysis by CTL, since highly potent CTL can be weakly efficient against allogeneic cells of different H-2 (refs 8 and 14). This could be due to the existence of different sub-populations of immune CTL, some of them reacting, for example, with antigens unrelated to H-2. (2) If HLA has the same role in humans as H-2 in mice, study of anti-tumour cell reaction due to CTL in cancer patients must be carried out in autologous conditions or, at least, between CTL and target cells sharing HLA determinants. Negative results would otherwise be of little significance.

Supported by contracts from INSERM and DGRST.

ELISABETH GOMARD
VERONIQUE DUPREZ
YVETTE HENIN
JEAN PAUL LEVY

Laboratoire d'Immunologie et virologie des tumeurs,
(Inserm U.136),
Service d'Hématologie Hôpital Cochin,
27 rue du Faubourg Saint Jacques,
75014 Paris, France

Received January 28; accepted March 9, 1976.

- ¹ Leclerc, J. C., Gomard, E., Plata, and Levy, J. P., *Int. J. Cancer*, **11**, 426-432 (1973).
- ² Leclerc, J. C., and Gomard, E., *Proc. 66th Am. Ass. Cancer Res.*, 806 (1975).
- ³ Levy, J. P., and Leclerc, J. C., *Adv. Cancer Res.* (in the press).
- ⁴ Zinkernagel, R. M., and Doherty, P. C., *J. exp. Med.*, **141**, 1427-1436 (1975).
- ⁵ Shearer, G. M., *Eur. J. Immun.*, **4**, 507-533 (1974).
- ⁶ Forman, J., *J. exp. Med.*, **142**, 403-417 (1975).
- ⁷ Ortiz de Landazuri, M., and Herberman, R. B., *Nature new Biol.*, **238**, 18-19 (1972).
- ⁸ Senik, A., Gisselbrecht, S., and Levy, J. P., *Int. J. Cancer*, **16**, 960-970 (1975).
- ⁹ Herberman, R. B., Nunn, M. E., Lavrin, D. M., and Asofsky, R., *J. natn. Cancer Inst.*, **51**, 1509-1512 (1973).
- ¹⁰ Bevan, M. J., *Nature*, **256**, 419-421 (1975).
- ¹¹ Koszinowski, U., and Ertl, M., *Nature*, **257**, 596-597 (1975).
- ¹² Gardner, I. D., Bower, N. A., and Blanden, R. N., *Eur. J. Immun.*, **5**, 122-127 (1975).
- ¹³ Blank, K. J., Freedman, H. A., and Lilly, F., *Nature*, **260**, 250-252 (1972).
- ¹⁴ Herberman, R. B., et al., *J. natn. Cancer Inst.*, **53**, 1103-1111 (1974).

Plasmin-mediated agglutination by concanavalin A of 3T3 cells cocultured with SV40-3T3 transformants

REICH and colleagues have reported that a variety of embryo fibroblasts in tissue culture show fibrinolytic activity after viral transformation^{1,2}. This effect is attributable to the action of a plasminogen activator, which is a serine protease³ and is present at higher concentrations in transformed than untransformed fibroblasts⁴; this activator converts plasminogen, present in the serum, to the fibrinolytic enzyme plasmin^{5,6}. Certain characteristics of transformed cells, such as altered morphology and multilayering, may be plasmin dependent^{7,8}. Plasmin may also induce changes in untransformed cells; this is supported by the observation that normal fibroblasts cocultured with viral transformants developed similar morphology to that of the transformed cells⁸. Furthermore, proteolytic activity released from transformed cells induced lectin-mediated agglutination, and stimulated thymidine incorporation in non-dividing cultures of untransformed fibroblasts and epithelial cells⁹. The possibility therefore exists that untransformed cells may express in coculture certain features of transformed cells as a result of environmental plasminogen activation. In support of this we report here the plasminogen-dependent enhancement of agglutination by concanavalin A (con A) of untransformed cells in coculture with transformed cells.

Swiss mouse 3T3 cells and SV40-3T3 transformants were used. We found that when 3T3 cells were grown in various sera they were generally only slightly agglutinable; however, when cocultured with SV40-3T3 cells they became highly agglutinable irrespective of the type of serum used (Table 1). Since contact between the two types of cell in coculture was only through the medium, this enhancement

Table 1 Agglutination by con A of 3T3 cells in monoculture and in coculture with SV40-3T3 cells

Serum supplement (10%)	Agglutination (monoculture)	Agglutination (coculture)	Net increase
Calf	1	8	7
Foetal calf	1	8	7
Horse	2	8	6
Pig	5	8	3
Chicken	1	6	5
Human	3	7	4
Guinea pig	3	7	4
Rabbit	1	7	6

To prevent mixing, the two cell types were initially seeded separately on microscope slides anchored down with silicone grease in square Petri dishes untreated for tissue culture. They were then incubated for 24 h in Dulbecco's modified Eagle's medium + 10% foetal calf serum. The slides of subconfluent cells were then thoroughly washed in serum-free medium, and transferred to new dishes for mono- or coculture in media supplemented with the sera listed. After 24 h they were washed twice and the cells were resuspended by incubating for 1 h in Ca²⁺- and Mg²⁺-free phosphate-buffered saline (PBS). Agglutination was then carried out in complete PBS + 100 µg ml⁻¹ con A in conical plastic tubes which were shaken at 100 r.p.m. for 30 min at 37 °C. Agglutination was then assessed on a 0-10 scale by 10 independent observers and an average unit value obtained.

is attributable either to the release of an SV40-3T3 cell effector substance or to the activation of a medium factor by an SV40-3T3 cell activator substance. The release of one or other of these agents into the medium was demonstrated when 3T3 cells incubated in serum-free medium conditioned by four 24-h exposures to SV40-3T3 cells showed enhanced agglutination in con A compared with those incubated in unconditioned medium (Table 2). The lower level of agglutination enhancement achieved in this case when compared with 3T3 cells cocultured for a single day in unconditioned medium (Table 1) suggests that the major part of the SV40-3T3 cell agent was cell associated. This tends to support the view that enhanced 3T3 cell agglutinability required the interaction of a cell activator substance with a component of the medium. We confirmed this by failing to enhance the agglutination of 3T3 cells in the absence of serum (Table 2). Consequently, the enhancement of 3T3 cell agglutination is attributable to the interaction of an activator of SV40-3T3 cell origin with a serum component. Plasminogen is a serum component and its activated form, plasmin, is a protease with trypsin-like specificity¹⁰. Since trypsin is known to induce con A agglutination of 3T3 cells¹¹ the enhanced agglutinability of 3T3 cells exposed to SV40-3T3 cell activator may be due to the conversion of serum plasminogen to plasmin. To test this we attempted to induce con A agglutination enhancement in cocultured 3T3 cells in serum-free medium to which plasminogen was added. After 24 h of coculture 3T3 cells became agglutinable, but cells incubated without plasminogen, SV40-3T3 cells, or both, failed to

Table 2 Agglutination by con A of 3T3 cells incubated in SV40-3T3 cell-conditioned medium

Serum supplement	Agglutination in unconditioned medium	Agglutination in conditioned medium	Net increase
10% calf	1	5	4
None	1	1	0

Serum-free Dulbecco's modified Eagle's medium was conditioned by exposure to SV40-3T3 cells for four separate 24-h periods, between each of which it was stored at -20 °C while the cells were allowed to recover for 24 h in foetal calf serum-supplemented medium. Before incubating with the 3T3 cells for 24 h the conditioned medium was filtered through 0.22 µm pores to remove any cells. 10% calf serum, selected for high agglutinating activity (Table 1), was added to the conditioned and unconditioned media. The subsequent incubation and agglutination procedures were as described in Table 1.

Table 3 Agglutination by con A of 3T3 cells in coculture with SV40-3T3 cells in serum-free medium containing plasminogen

Cocultured with SV40-3T3 cells	Medium supplement	Agglutination
+	Plasminogen	4
—	Plasminogen	1
+	None	1
—	None	1
+	Plasminogen	1 (no con A added)
—	10% foetal calf serum	1

Cells were grown on microscope slides as described in Table 1. After washing they were incubated for 24 h in the conditions shown; lyophilised pig blood plasminogen was added at $20 \mu\text{g ml}^{-1}$. Since cells incubated without serum were more readily detached from the substrate than those incubated with serum, the incubation in Ca^{2+} - and Mg^{2+} -free PBS was reduced to 10 min in these groups. Agglutination was carried out as described in Table 1.

agglutinate (Table 3). Consequently, enhanced agglutination is attributable to the interaction of SV40-3T3 cell activator and plasminogen. Using the casein-Agarose overlay assay⁴, we found that detectable digestion of casein was produced by SV40-3T3 cells, but not by 3T3 cells, in media supplemented with any of the sera listed in Table 1. This supports the view that the SV40-3T3 cell factor is a plasminogen activator which converts added or serum plasminogen to plasmin, and that it is present at lower activity in, or absent from, 3T3 cells.

It has been reported¹² that human uterine cervix fibroblasts, normally not agglutinable with con A, show increasing agglutination as adjacent lesions of the overlying epithelium progress from dysplasia through *in situ* carcinoma to invasive carcinoma. There may be a close parallel between these changes and the plasmin-mediated enhancement of agglutination in con A of cocultured 3T3 cells reported here. Since primary cultures of malignant tissues may often have high concentrations of plasminogen activator compared with normal tissues¹³, the increasing agglutination of normal fibroblasts adjacent to these neoplastic lesions may be attributable to a local conversion of serum plasminogen into plasmin by increasing concentrations of tumour cell plasminogen activator. This suggests the interesting possibility that tumours may be associated with normal cells in a "quasi-malignant" state, expressing one or more features usually considered to be characteristic of malignant transformation.

PAUL WHUR
HANNO KOPPEL
CATRIONA URQUHART
DONALD C. WILLIAMS

Marie Curie Memorial Foundation,
Oxted, Surrey, UK

Received January 23; accepted February 24, 1976.

- Unkles, J. C., Tobia, A., Ossowski, L., Quigley, J. P., Rifkin, D. B., and Reich, E., *J. exp. Med.*, **137**, 85-111 (1973).
- Ossowski, L., Unkles, J. C., Tobia, A., Quigley, J. P., Rifkin, D. B., and Reich, E., *J. exp. Med.*, **137**, 112-126 (1973).
- Christman, J. K., and Acs, G., *Biochim. biophys. Acta*, **340**, 339-347 (1974).
- Goldberg, A. R., *Cell*, **2**, 95-102 (1974).
- Unkles, J., Danø, K., Kellerman, G. M., and Reich, E., *J. biol. Chem.*, **249**, 4295-4305 (1974).
- Quigley, J. P., Ossowski, L., and Reich, E., *J. biol. Chem.*, **249**, 4306-4311 (1974).
- Ossowski, L., Quigley, J. P., Kellerman, G. M., and Reich, E., *J. exp. Med.*, **138**, 1056-1064 (1973).
- Ossowski, L., Quigley, J. P., and Reich, E., *J. biol. Chem.*, **249**, 4312-4320 (1974).
- Goldberg, A. R., *Abstr. Ann. Mtg Am. Soc. Microbiol.*, **218** (1973).
- Robbins, K. C., and Summaria, L., in *Methods in Enzymology*, **19**, 184-199 (Academic, New York, 1970).
- Rottmann, W. L., Walther, B. T., Hellerqvist, C. G., Umbreit, J., and Roseman, S., *J. biol. Chem.*, **249**, 373-380 (1974).
- Chaudhuri, S., Koprowska, I., and Rowinski, J., *Cancer Res.*, **35**, 2350-2354 (1975).
- Laug, W. E., Jones, P. A., and Benedict, W. F., *J. natn. Cancer Inst.*, **54**, 173-178 (1975).

Transformation of C3H/10T1/2 CL8 mouse embryo fibroblasts by ultraviolet irradiation and a phorbol ester

ULTRAVIOLET light is the major cause of human skin cancer, and papillomas and malignant carcinomas have been induced in the skins of rats and mice by repeated exposure to it¹⁻⁵. Pound⁶ reported that although a single short exposure of mouse skin to ultraviolet light did not induce any tumours, when the single exposure was followed by repeated applications of croton oil, many tumours were induced. Thus ultraviolet light acted as an initiator in the two-stage process of skin tumour development^{7,8}. X rays can induce oncogenic transformation of cells in culture^{9,10}, and a preliminary report has appeared on the transformation of hamster embryo cells by ultraviolet light¹¹.

We have developed a cloned line of C3H mouse fibroblasts, C3H/10T1/2 CL8, which is non-malignant and highly susceptible to postconfluence inhibition of cell division¹². These cells are transformed to malignancy with polycyclic hydrocarbon carcinogens and by certain alkylating agents¹³, and the response can be quantified. We have also reported¹⁴ that when these cells were treated with a subeffective concentration of polycyclic hydrocarbons followed 5 d later by a non-transforming and non-cytotoxic concentration of tetradecanoyl phorbol acetate (TPA, $0.1 \mu\text{g ml}^{-1}$)¹⁵, transformation occurred. Thus, initiation and promotion have been achieved with our system of cultured cells. The transformed foci so obtained are of type III, (ref. 13) which give rise to fibrosarcomas in syngeneic mice¹³.

In our experiments, 2,000 C3H/10T1/2 cells were plated in 60-mm Falcon dishes in Eagle's BME medium + 10% heat-inactivated foetal calf serum (Gibco). Twenty-four hours later the cells were irradiated once in fresh medium with an ultraviolet lamp (Dazor) with doses from 10 to 100 ergs mm^{-2} . The dose of ultraviolet irradiation was measured by an ergometer (Ultra-Violet Products) in $\text{ergs mm}^{-2} \text{ s}^{-1}$. After irradiation the cells were given fresh medium and incubated in a humidified incubator (Forma) at 37°C in an atmosphere of 5% carbon dioxide in air. After various times some of the cultures were given TPA ($0.1 \mu\text{g ml}^{-1}$) in the medium for the remainder of the experiment. TPA (Consolidated Midland) was dissolved in spectrograde acetone and stored in amber bottles at -20°C . It was added to the medium so that the final concentration of acetone was 0.5%. The cytotoxicity of the ultraviolet irradiation was determined by its effect on the plating efficiency of 200 cells, and the colonies were

Fig. 1 Fixed and stained dishes of C3H/10T1/2 CL8 cells 6 weeks after initial treatment: 1, 0.5% acetone; 2, TPA $0.1 \mu\text{g ml}^{-1}$; 3, ultraviolet light, 25 erg mm^{-2} ; 4, ultraviolet light, 25 erg mm^{-2} , followed 48 h later by TPA $0.1 \mu\text{g ml}^{-1}$.

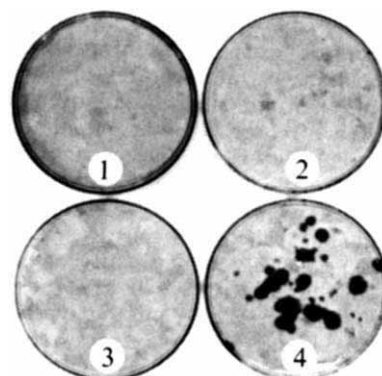


Table 1 Transformation of C3H/10T1/2 CL8 cells by ultraviolet light and TPA

Treatment schedule	Plating efficiency %	No. of dishes with foci/total no. of dishes	% Dishes with foci	Transformation frequency*
0.5% Acetone	19	0/23	0	0
TPA (0.1 µg ml ⁻¹)	20	0/12	0	0
Ultraviolet, 10 erg mm ⁻²	19	0/12	0	0
Ultraviolet, 10 erg mm ⁻² +TPA				
48 h†	19	3/12	25	0.08
72 h	19	5/12	42	0.18
96 h	19	4/12	33	0.13
120 h	19	3/12	25	0.11
Ultraviolet, 25 erg mm ⁻²	17	0/16	0	0
Ultraviolet, 25 erg mm ⁻² +TPA				
48 h	18	10/12	83	0.83
72 h	18	11/12	92	0.69
96 h	18	10/12	83	1.0
120 h	18	9/12	75	0.35
Ultraviolet, 50 erg mm ⁻²	8	0/24	0	0
Ultraviolet, 50 erg mm ⁻² +TPA				
48 h	7	8/12	67	0.95
72 h	7	8/12	67	1.6
96 h	7	7/12	58	1.9
120 h	7	5/11	45	0.45
Ultraviolet, 100 erg mm ⁻²	1	0/18	0	0
Ultraviolet, 100 erg mm ⁻² +TPA				
48 h	1	6/15	40	3.7
72 h	1	8/17	47	2.6
96 h	1	3/12	25	1.7
120 h	1	3/11	27	1.4

The concentration of TPA used was 0.1 µg ml⁻¹.

* Percentage transformed foci per surviving cell plated.

† Time after irradiation when TPA was added.

counted after 8–10 d (ref. 13). Type III transformed foci were scored 6 d after fixation with methanol and staining with Giemsa, which was 6 weeks after 2,000 cells had been plated in 60-mm dishes, as described previously^{10,13}.

The results are shown in Table 1. Ultraviolet irradiation at 10 and 25 erg mm⁻² was not cytotoxic, but larger doses were. No type III transformed foci were obtained in the acetone control dishes, in the dishes treated with TPA (1 µg ml⁻¹) or in the dishes treated with ultraviolet irradiation. When TPA was added to the medium 48 h or more after irradiation, however, significant numbers of transformed foci were observed (Table 1 and Fig. 1). This transformation was dose dependent for ultraviolet irradiation, with high transformation frequencies, corrected for cytotoxicity, for 100 erg mm⁻² followed by TPA. In general, the transformation frequency decreased if TPA treatment started 96 h or more after irradiation.

These experiments show that ultraviolet light acts as an initiator of transformation in our cells, just as it did in the experiments on mouse skin⁶. This first success in achieving transformation in cell cultures with ultraviolet light will lead to important information on the mechanism of ultraviolet carcinogenesis. Relevant studies, with particular emphasis on DNA repair, are in progress.

We thank Mr D. W. Brankow for technical assistance. This work was supported by a contract from the National Cancer Institute. C. H. is an American Cancer Society Professor of Oncology.

SUKDEB MONDAL
CHARLES HEIDELBERGER

McArdle Laboratory for Cancer Research,
University of Wisconsin,
Madison, Wisconsin 53706

Received January 5; accepted February 24, 1976.

¹ Blum, H. F., *Carcinogenesis by Ultraviolet Light* (Princeton University Press, Princeton, 1959).

² Findlay, G. M., *Lancet*, ii, 106, 1070–1073 (1928).

³ Findlay, G. M., *Lancet*, i, 108, 1229–1231 (1930).

⁴ Funding, G., Henriques, O. M., and Reckling, E., *Dritter Int. Kong. Lichtforschung*, Wiesbaden, 166 (1936).

⁵ Rusch, H. P., Kline, B. E., and Baumann, C. A., *Archs Pathol.*, 31, 135–146 (1941).

⁶ Pound, A. W., *Pathology*, 2, 269–275 (1970).

⁷ Berenblum, I., *Cancer Res.*, 1, 44–48 (1941); *ibid.*, 14, 471–477 (1954).

⁸ Boutwell, R. K., *Prog. exp. Tumor Res.*, 4, 207–250 (1964); *CRC Crit. Revs Toxicol.*, 419–443 (1974).

⁹ Borek, C., and Sachs, L., *Nature*, 210, 276–278 (1966).

¹⁰ Terzaghi, M., and Little, J. B., *Nature*, 253, 548–549 (1975).

¹¹ Di Paolo, J. A., and Donovan, P. J., *Proc. Am. Ass. Cancer Res.*, 16, 75 (1975).

¹² Reznikoff, C. A., Brankow, D. W., and Heidelberger, C., *Cancer Res.*, 33, 3231–3238 (1973).

¹³ Reznikoff, C. A., Bertram, J. S., Brankow, D. W., and Heidelberger, C., *Cancer Res.*, 33, 3239–3249 (1973).

¹⁴ Mondal, S., Peterson, A. R., and Heidelberger, C., *Proc. Am. Ass. Cancer Res.*, 16, 74 (1975).

¹⁵ Hecker, E., *Meth. Cancer Res.*, 6, 439–484 (1971).

Increased procoagulant activity in cultured fibroblasts from progeria and Werner's syndromes of premature ageing

BIOLOGICAL ageing undoubtedly involves a number of complex physiological and biochemical factors but it is increasingly evident that the cultured human fibroblast will help to elucidate the fundamental nature of this process. These cells have a finite replicative lifespan that is inversely related to the age of the donor^{1–3}. Moreover, physiological rather than chronological age is a critical determinant, since cultured fibroblasts from the progeria and Werner's syndromes of premature ageing have significantly reduced growth potential compared with normal, age-matched controls^{3–5}. Although the decreased replicative ability of cells may help to explain some of the features of progeria and Werner's syndrome, such as stunting of growth and impaired wound healing ability, it does not explain other pathological concomitants of these disorders. In particular, an explanation is needed for the premature appearance of severe atherothrombotic disease^{6,7}. It has been shown that human fibroblasts contain a potent procoagulant activity called 'tissue factor' (TF)⁸. TF has the capacity to activate the extrinsic clotting mechanism through factor VII, eventually leading to thrombin generation which then promotes clot formation through fibrinogen^{8,9}. We report here that skin fibroblasts derived from subjects with progeria and

Werner's syndrome have markedly elevated levels of TF compared with controls.

Fibroblast strains were initiated from skin biopsies of three male subjects with classical progeria, one aged 2.5 yr and two aged 9 yr, and a 56-yr-old woman with Werner's syndrome, following informed consent of the subjects or their parents. Clinical details on two of the progeria subjects¹⁰ and the subject with Werner's syndrome^{3,6} have been published. Normal control cultures were derived from donors aged 22–76 yr, whereas abnormal controls were two subjects aged 5 yr with Lesch–Nyhan syndrome and homocystinuria. The latter two were better matched to the progeria subjects in terms of age, stunting of growth and chronic illness, all of which could possibly affect the growth of cultured cells and the amount of TF activity. All cultures were initiated and propagated in our laboratory using Eagle's medium supplemented with non-essential amino acids and 15% foetal calf serum as described previously^{2,10}. Subcultivation was carried out at a 1:8 split ratio counting three mean population doublings each time, until exhaustion of the replicative potential^{1–3}. By this criterion, control cells had replicative lifespans ranging from 48–75 doublings, progeria cells 21–40 doublings, and Werner's strain only 15 doublings. Control cells were assayed for TF both at early and late passage, the latter 3–9 doublings before termination. Progeria cells were assayed with 9–30 doublings remaining and Werner's cells within 3–6 doublings of termination. In preparation for assay of TF activity, post-confluent sheets of cells were collected by a routine trypsinisation method^{1–3} that was strictly standardised. Cells were incubated with a 0.125 trypsin solution (Difco 1:250) at 37 °C for exactly 10 min, after which the trypsin was neutralised with a 20-volume excess of growth medium containing 15% foetal calf serum. When soybean trypsin inhibitor was used to neutralise trypsin, essentially identical results were obtained in the TF assay.

Figure 1 is a combined plot of TF activities in the various fibroblast strains. TF in control cells ranged from 2 to 12 U per 10⁶ cells with a mean of $5.06 \text{ U} \pm 3.84$ (s.d.). No effect of donor age was evident. In striking contrast, progeria and Werner's cells showed significantly elevated TF activities, ranging from 23 U to a maximum of 137 U. Since the anatomic site of biopsy can influence various parameters in cultured fibroblasts³, assays on cell strains derived from skin of the anterior forearm of the 2.5- and 9-yr-old progeria subjects ($n=16$) were compared with controls, all derived from the same anatomic site ($n=41$). The differences remained highly significant ($P<0.01$). Slow growth *per se* of the prematurely ageing cells could not account for these findings; when control cells were assayed during their senescent phase, with 3–9 mean population doublings remaining until termination of growth, the mean TF activity was $7.96 \pm 6.82 \text{ U}$, still significantly different from progeria strains ($P<0.01$). Additionally, control strains from aged adults, with growth properties similar to progeria cells¹⁰ had substantially lower TF activities (Fig. 1). The apparent increment in TF activity during the lifespan of control cells was not statistically significant.

Disruption of cells by sonication produced about a five-

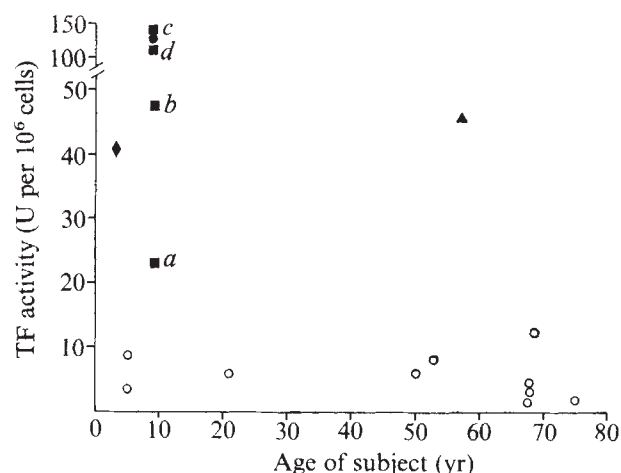


Fig. 1 TF activity of early passage fibroblasts as a function of the age of donor. The assay was slightly modified from the one-stage thromboplastin method of Zacharski and McIntyre⁸. Collected cells were suspended and rinsed twice in PBS and then in Tyrode's albumin buffer¹¹. The final suspension of washed fibroblasts was adjusted to 2×10^6 cells ml^{-1} . Platelet-poor citrated human plasma obtained from healthy donors and the plasmas of patients with hereditary factor VII and factor VIII deficiency (Dade Division, American Hospital Supply Corporation, Miami, Florida 33152) were used as substrates. To carry out the clotting assay, 0.1-ml aliquots of plasma were incubated in a water bath at 37 °C, followed by addition of 0.1 ml fibroblast suspension and 0.1 ml 0.025 M CaCl_2 . The clotting times were recorded and the procoagulant activity of fibroblasts was compared with that of rabbit brain thromboplastin (Thromboplastin reagent, Dade). We accepted arbitrarily that 1 ml of this reagent contained 1,000 U TF. Thromboplastin was serially diluted up to 1:5,120 in Tyrode's albumin buffer and the effect of each dilution was tested on the clotting time. The clotting time of normal human plasma without rabbit brain thromboplastin was approximately 210 s, whereas the sample with concentrated thromboplastin clotted in 14–15 s. Since dilution curves obtained using rabbit brain thromboplastin and fibroblasts were of a similar character, TF activity of fibroblasts was expressed in arbitrary units calculated by interpolation. Each fibroblast suspension was assayed at a minimum of four cell concentrations. Progeria (solid symbols), Werner's syndrome $\blacktriangle$, and normal controls (open symbols). *a–d*, Single subject with progeria using individual cultures initiated from skin of various anatomic sites: *a*, anterior forearm; *b*, anterior thigh; *c* and *d*, bilateral calves. Strain $\bullet$ originated from skin over the lumbar area, whereas all other cultures originated from skin of the anterior forearm. Values are means of 3–14 experiments except for *b*, *c* and *d*, which represent single observations.

fold increase in TF activity in all strains but the differences between controls and prematurely ageing cells were maintained. This suggests that the higher levels of TF in progeria and Werner's cells relate primarily to increased content rather than increased availability. The supernatant obtained following 300g centrifugation of cells collected by trypsinisation and neutralised with soybean trypsin inhibitor was also assayed, but no TF activity could be detected. In contrast, control supernatants, obtained after incubating monolayers with enzyme-free buffer, contained about 10% of the total TF in the system (cells plus supernatant). These results

Table 1 Effect of cultured skin fibroblasts on clotting time of recalcified plasma in presence of normal and factor-deficient plasma

Additive	Normal		Clotting time (s) Type of plasma		Factor VIII-deficient	
	Expt 1	Expt 2	Factor VII-deficient Expt 1	Expt 2	Expt 1	Expt 2
Control (Tyrode's fluid alone)	186	166	227	154	527	510
Normal cells	49	80	176	145	58	61
Progeria cells	31	53	120	123	36	31

Early passage cells from 22-yr-old normal male and 9-yr-old boy with progeria were used.

are in conflict with a report using more purified trypsin¹² and suggest that different experimental techniques may be responsible. For example, the relatively crude pancreatic trypsin used in our studies may contain other enzymes such as lipases capable of digesting an essential lipid component of TF^{12,13}.

To determine whether fibroblast TF was specific for the extrinsic clotting mechanism (that is, activation via factor VII), experiments were carried out with intact cells using plasma from patients with hereditary deficiencies of clotting factors (Table 1). With factor VII-deficient plasma, both normal and progeric cells showed a markedly prolonged clotting time, much like the control fluid without cells. When factor VIII-deficient plasma was used, however, the clotting time was again shortened to resemble values observed with normal plasma. In total, these data strongly suggest that the extrinsic rather than the intrinsic clotting mechanism plays an important role in the TF-mediated clotting.

To determine whether TF itself was responsible for the increased procoagulant activity, cells were preincubated with specific immune globulins directed against purified TF followed by assay (Fig. 2). Both normal and progeria cells showed a progressive prolongation of the clotting time after treatment with the specific immune globulins, but not with preimmune globulins or normal saline. This indicates that fibroblast TF is similar or identical to the placental TF used as antigen, and further, that both can stimulate clotting through the extrinsic mechanism.

It is not clear why cells from two genetically distinct disorders^{6,7} show similar elevations in TF activity. Reports indicate that progeria and Werner's cells contain a diversity of altered proteins affecting HLA antigens, cytoplasmic enzymes^{10,14-16} and insulin receptors²³. Thus, although the TF of prematurely ageing cells seems to be recognised normally by specific antiserum, it may have a structural alteration such that a given amount possesses increased procoagulant activity.

An explanation is also needed for how an increased amount of procoagulant in prematurely ageing skin fibroblasts could relate to the severe vascular disease of progeria and Werner's syndrome^{6,7}. Indeed, it would be extremely

useful to know the TF activity and growth potential of fibroblasts from a second set of controls, subjects with severe atherothrombotic disease apart from progeria and Werner's syndromes. Such studies, now in progress, may confirm that increased procoagulant activity and impaired growth potential are important factors in atherothrombosis when they involve certain cells within the vascular tree¹⁴. In fact, there is evidence that the cultured 'fibroblast' may originate from vascular cells such as endothelium or pericytes^{17,18}. The early concept of Rokitansky¹⁹, still extant today^{20,21}, proposes that coagulation mechanisms acting in or on arterial walls contribute to atherosclerosis. Zeldis *et al.*²² have found that plasma membranes of intimal and medial cells are rich in TF antigen, in large blood vessels and their vasa vasorum. Moreover, TF is particularly abundant in the vicinity of atherosclerotic plaques where it virtually encompasses cholesterol crystals. Thus, TF could initiate coagulation not only as a consequence of plaque disruption but also after minimal cellular damage^{12,22}. In any case, it is possible that elevated TF in arterial cells is an important factor that predisposes individuals with progeria and Werner's syndromes to an increased risk of thrombosis, atherogenesis and premature death.

We thank C. Lowery and E. Moerman for assistance and Dr Y. Nemerson for antiserum to TF. Supported by the NIH the MRC of Canada, the Canadian Diabetic Association Foundation Fund and the Ontario Heart Fund during the tenure of a MRC Scholarship to S.G.

SAMUEL GOLDSTEIN*

Departments of Medicine and Biochemistry,
McMaster University Medical Centre,
Hamilton, Ontario, Canada

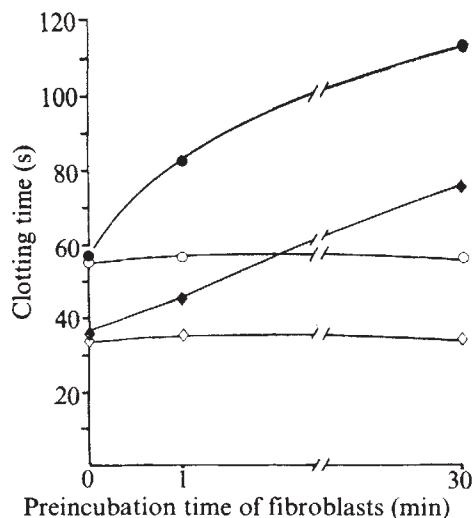
STEFAN NIEWIAROWSKI

Specialized Center for Thrombosis Research,
and Department of Medicine,
Temple University,
Philadelphia, Pennsylvania

Received October 27, 1975; accepted March 11, 1976.

* Present address: National Institute for Medical Research, Mill Hill, London NW7 1AA, UK.

Fig. 2 TF activity of cultured fibroblasts incubated with and without specific immune globulins against purified TF. Fibroblasts were preincubated with immune globulins, preimmune globulins, or 0.9 NaCl and then assayed for TF. Early passage cells from a 22-yr-old normal male and from the 2.5-yr-old boy with progeria were used. ●, Normal cells+immune globulins; ◆, progeria cells+immune globulins; ○, normal cells+0.9% NaCl or preimmune globulins; ◇, progeria cells+0.9% NaCl or preimmune globulins.



- Hayflick, L., *Expl Cell Res.*, **37**, 614 (1965).
- Goldstein, S., Littlefield, J. W., and Soeldner, J. S., *Proc. natn. Acad. Sci. U.S.A.*, **64**, 155 (1969).
- Martin, G. M., Sprague, C. A., and Epstein, C. J., *Lab. Invest.*, **23**, 86 (1970).
- Goldstein, S., *Lancet*, **i**, 424 (1969).
- Danes, B. S., *J. clin. Invest.*, **50**, 2000 (1971).
- Epstein, C. J., Martin, G. M., Schultz, A. L., and Motulsky, A. G., *Medicine*, **45**, 177 (1966).
- DeBusk, F. L., *J. Pediatr.*, **80**, 697 (1972).
- Zacharski, L. R., and McIntyre, O. R., *Blood*, **41**, 679 (1973).
- Davie, E. W., and Fujikawa, K., *A. Rev. Biochem.*, **44**, 779 (1975).
- Goldstein, S., and Moerman, E., *New Engl. J. Med.*, **292**, 1306 (1975).
- Niewiarowski, S., and Goldstein, S., *J. Lab. clin. Med.*, **82**, 605 (1973).
- Maynard, J. R., Heckman, C. A., Pitlick, F. A., and Nemerson, Y., *J. clin. Invest.*, **55**, 814 (1975).
- Nemerson, Y., and Pitlick, F. A., *Biochemistry*, **9**, 5100 (1970).
- Goldstein, S., Niewiarowski, S., and Singal, D. P., *Fedn Proc.*, **34**, 56 (1975).
- Goldstein, S., and Singal, D. P., *Nature*, **251**, 719 (1974).
- Goldstein, S., and Moerman, E. J., *Nature*, **255**, 159 (1975).
- Papayannopoulou, T. G., and Martin, G. M., *Expl Cell Res.*, **45**, 72 (1966).
- Franks, L. M., and Couper, T. W., *Int. J. Cancer*, **9**, 19 (1972).
- Rokitansky, C., *A Manual of Pathological Anatomy*, **4** (translated by Day, G. E.) (Sydenham Society, London, 1852).
- Duguid, J. B., *Br. med. Bull.*, **11**, 36 (1955).
- Moore, S., *Thrombos. Diathes. Haem.*, **Stuttg.**, **33**, 417 (1975).
- Zeldis, S. M., Nemerson, Y., and Pitlick, F. A., *Science*, **175**, 766 (1972).
- Rosenbloom, A., Goldstein, S., and Yip, C., *Science* (in the press).

Thymic factor in human sera demonstrable by a cyclic AMP assay

NORMAL human serum has been reported to contain thymus hormone activity as detectable by a rosette inhibition assay¹. It has been suggested that this circulating hormone may act through stimulation of cyclic AMP synthesis². In addition, other thymic hormones isolated from the thymus, such as thymosin³ and THF⁴, have been found to activate membrane adenylate cyclase and to increase cellular levels of cyclic AMP in thymocytes^{5,6}. Thymopoietin, a thymic hormone

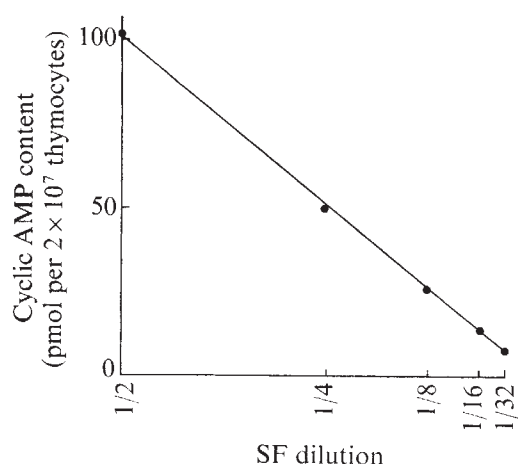


Fig. 1 Effect of different dilutions of SF on cellular cyclic AMP level. Human serum was collected by centrifugation at 4 °C after defibrination of the blood. Serum was filtered on Amicon membrane CF50 (nominal cut-off molecular weight 50,000) by which high molecular weight thymic hormone inhibitors were removed⁹. Ultrafiltrate was added at different dilutions to Earle's solution containing thymocytes obtained from C57BL/6 mice aged 4–8 weeks. Cell concentration was adjusted to 2×10^6 cells ml⁻¹. After 5 min of incubation at 37 °C cellular cyclic AMP was extracted, placing tubes in liquid nitrogen for 1 min and subsequently in a boiling water bath for 5 min. Coagulated proteins were spun down in the cold and the supernatant directly assayed in duplicate, by Gilman's competition-binding procedure¹⁰ with a cyclic AMP assay kit (TRK432, The Radiochemical Centre).

the amino acid sequence of which has been established⁷, has also been claimed to induce T-cell differentiation by way of cyclic AMP⁸. Here we report on a human serum factor (SF) capable of increasing the level of cellular cyclic AMP in thymocytes; we also present evidence for the possible thymic nature of this factor.

Figure 1 shows the effect of SF from a normal 18-yr-old

Table 1 Effect of adsorptions on SF activity

	pmol per 2×10^7 thymocytes
No SF added	12.8 ± 0.8
SF (dilution 1:4)	61.8 ± 2.5
SF (1:4) adsorbed on human thymocytes	11.0 ± 1.4
SF (1:4) adsorbed on mice thymocytes	18.2 ± 0.5
SF (1:4) adsorbed on human granulocytes	61.8 ± 7.6
SF (1:4) adsorbed on mouse liver cells	88.4 ± 0.6

Results are expressed as mean ± s.e. of four experiments. SF was incubated for 1 h at 37 °C either alone or in the presence of 1×10^6 lyophilised human thymocytes or granulocytes per ml SF. When viable mouse thymocytes and liver cells were used for adsorption, the ratio of 4×10^6 cells per ml SF was used.

donor on cyclic AMP levels in mouse thymocytes. Similar results were obtained with human thymocytes. Among sera from different donors a marked variability in SF activity was observed. The differences in the increase of cyclic AMP levels were, however, in no way correlated with the cyclic AMP content of the sera (correlation coefficient=0.0304,

$n=19$). In an attempt to establish the thymic origin of the SF we investigated sera from a total of 18 thymectomised myasthenic patients (mean age 35.6 yr, s.d. 13.8). The SF activity was virtually absent in the thymectomised donors (mean 12.64 pmol cyclic AMP/ 2×10^7 cells, s.e. 1.96), whereas in normal healthy individuals ($n=399$; mean age 26.7 yr, s.d. 7.8) and in non-thymectomised myasthenic patients ($n=5$; mean age 47 yr, s.d. 13.3) was consistently demonstrable (mean 76.69 pmol cyclic AMP/ 2×10^7 cells, s.e. 6.2; and mean 92.80 pmol cyclic AMP/ 2×10^7 cells, s.e. 21.81, respectively). When SF from normal donors was mixed with a similarly prepared material from thymectomised, no decrease in the activity was observed. The presence of SF inhibitors in the SF from thymectomised was therefore ruled out.

We also investigated the binding capacity of SF to different cell types. To this end, adsorption experiments were carried out. It was found that SF activity could be completely adsorbed by lyophilised human thymocytes as well as by viable mouse thymocytes. On the other hand, adsorption with either lyophilised human granulocytes or viable mouse liver cells did not lead to a diminution of SF activity (Table 1).

Finally, the target specificity of SF and a number of agents known to elevate cellular cyclic AMP were tested on a panel of different mouse cells (Table 2). It seems therefore that SF is increasing cyclic AMP levels mainly in thymocytes with little or no effect on other lymphoid cells. In contrast, nonspecific stimulators of cyclic AMP, such as isoproterenol and prostaglandin E₁ (PGE₁), showed a marked stimulation in all target cells tested. Note that the increase in cyclic AMP in thymocytes given by the β -adrenergic activator isoproterenol could be selectively inhibited by the antagonist propranolol. Since the effect of SF is not inhibited to any extent by propranolol (Table 3), it seems that

Table 3 Effects of SF, isoproterenol and propranolol on cyclic AMP level in thymocytes

	pmol per 2×10^7 thymocytes
No SF added	12.4 ± 0.6
SF (dilution 1:4)	72.3 ± 3.5
Isoproterenol (10^{-5} M)	220.9 ± 1.0
Propranolol (10^{-5} M)	11.7 ± 0.8
SF (1:4) + isoproterenol (10^{-5} M)	236.7 ± 16.2
SF (1:4) + propranolol (10^{-5} M)	76.1 ± 8.1
Isoproterenol (10^{-5} M) + propranolol (10^{-5} M)	10.5 ± 0.1
SF (1:4) + isoproterenol (10^{-5} M) + propranolol (10^{-5} M)	77.5 ± 2.3

Results expressed as mean ± s.e. of four experiments.

SF is acting through a membrane site, probably a receptor distinct from that for β -adrenergic activators. When isoproterenol and SF were added together, only a partial additive stimulation occurred. This indicates that, although they act on distinct membrane sites, isoproterenol and SF may stimulate the same adenylate cyclase regulatory subunit, as proposed for thymic hormones⁸. The combination isoproterenol, propranolol and SF produced a similar increase in cyclic AMP as did SF alone, again indicating that the antagonism isoproterenol–propranolol did not interfere with

Table 2 Effects of SF and drugs on cyclic AMP levels in different target cells

	Thymocytes	Spleen cells	Lymph node cells	PBL*	Liver cells
No SF added	10.3 ± 0.3	11.7 ± 1.4	23.0 ± 0.2	37.2 ± 1.3	15.3 ± 0.9
SF (dilution 1:2)	139.1 ± 1.3	17.2 ± 1.0	56.7 ± 4.0	28.8 ± 1.7	13.7 ± 0.8
PGE ₁ (10^{-5} M)	151.8 ± 9.9	42.7 ± 6.8	107.0 ± 3.7	113.1 ± 1.6	47.0 ± 0.8
Isoproterenol (10^{-5} M)	218.8 ± 1.1	46.5 ± 9.3	83.1 ± 2.0	81.4 ± 1.7	37.4 ± 1.3

Results expressed in pmol per 2×10^7 cells (mean ± s.e. of four experiments).

*Human peripheral blood lymphocytes.

the capacity of the cells to respond to SF by an increased cyclic AMP level.

In conclusion, SF seems to be thymus derived, since the activity in thymectomised patients is absent. Furthermore, among the different target cells investigated, only thymocytes were found to possess a membrane site for SF in addition to the β -adrenergic receptor. Experiments are now being carried out to establish whether SF has the biological activity of the thymic hormone described by Bach *et al.*¹. The stimulation of cyclic AMP synthesis on thymocytes will eventually provide a quantitative and reproducible assay for the detection of human circulating thymic hormone.

A.A. is a European Molecular Biology Organization long-term fellow.

A. ASTALDI
GIULIA C. B. ASTALDI
P. TH. A. SCHELLEKENS
V. P. EIJSSVOOGEL

Central Laboratory of the Netherlands
Red Cross Blood Transfusion Service,
Laboratory of Experimental and Clinical Immunology
of the University of Amsterdam,
POB 9190, Amsterdam, The Netherlands

Received January 12; accepted March 5, 1976.

¹ Bach, J. F., *et al.*, *Lancet*, ii, 1056–1058 (1972).

² Bach, M. A., Fournier, C., and Bach, J. F., *Ann. N.Y. Acad. Sci.*, **249**, 316–327 (1975).

³ Hooper, J. A., *et al.*, *Ann. N.Y. Acad. Sci.*, **249**, 125–144 (1975).

⁴ Kook, A. I., Yakiz, Y., and Trainin, N., *Cell. Immun.*, **19**, 151–157 (1975).

⁵ Kook, A. I., and Trainin, N., *J. exp. Med.*, **139**, 193–207 (1974).

⁶ Kruisbeek, A. M., in *The Biological Activity of Thymic Hormones* (edit. by van Bekkum, D. W.), 209–211 (Kooiker Scientific, Rotterdam, 1975).

⁷ Schlesinger, D. H., and Goldstein, G., *Cell*, **5**, 361–365 (1975).

⁸ Scheid, M. P., Goldstein, G., Hammerling, U., and Boyse, E. A., *Ann. N.Y. Acad. Sci.*, **249**, 531–538 (1975).

⁹ Bach, J. F., and Dardenne, M., *Immunology*, **25**, 353–366 (1973).

¹⁰ Gilman, A. G., *Proc. natn. Acad. Sci. U.S.A.*, **67**, 305–312 (1970).

Lack of an autosomal recessive genetic influence in vertical transmission of hepatitis B antigen

BLUMBERG *et al.* have postulated that predisposition to the hepatitis B antigen (HBsAg) carrier state is inherited as an autosomal recessive gene¹. Crucial to this concept, but lacking in Blumberg's populations, are families in which both parents are carriers. All offspring of such families should be homozygous for carrier predisposition and once infected all should become carriers. To test Blumberg's hypothesis, we compared rates of hepatitis B antigenaemia among children in two groups of families: those in which both parents were HBsAg carriers and those in which the mother was a carrier, but the father's serum was antibody (anti-HBs) positive. We did not include families when the father's serum was both HBsAg and anti-HBs negative, since carrier state predisposition remained unknown without evidence of previous hepatitis B virus (HBV) infection. If the autosomal recessive genetic hypothesis is correct, we would expect, in families with two HBsAg carrier parents, that all HBV-infected offspring would become carriers. In contrast, only a small proportion of children would become carriers in families with a carrier mother and an antibody-positive father, since most children would have only a single gene for carrier predisposition (inherited from their mothers) and would, like their fathers, form antibody after exposure rather than becoming carriers.

The families in this study were derived from a study of

vertical transmission of the hepatitis B virus from carrier mothers in Taiwan to their newborn offspring. The details of the population studied and methods used have been published before^{2,3}. Sera of newborns were tested periodically for HBsAg and anti-HBs, and other family members' sera were tested at least once. We have included here all families in which the mothers' sera contained HBsAg and the fathers' sera contained either HBsAg or anti-HBs. Data are included only on children who were at least 3 months old (when most babies who became carriers were first known to be HBsAg positive)².

Hepatitis B surface antigen was detected in all specimens by radioimmunoassay (RIA, AUSRIA-ITM, Abbott Diagnostics)⁴ and complement fixation (CF)^{5,6}. Babies' sera were retested by the more sensitive AUSRIA-IITM, so rates of antigenaemia among the babies are higher than described previously². Antibody (anti-HBs) was detected by passive haemagglutination (PHA-Electronucleonics) with titres of 1 : 8 or greater considered positive⁷.

The frequency of antigenaemia among the offspring (both the babies and their siblings) was unrelated to whether their father's serum was antigen or antibody positive (Table 1). Babies in both groups were followed from birth for the same time (mean, 12 months) and none of the babies or their siblings had clinical hepatitis. All but five of 27 babies retested after their first positive specimens (eight of nine with an antigen-carrier father and 14 of 18 with an antibody-positive father) remained HBsAg carriers (average of 11 months follow-up). One further baby in each group developed anti-HBs (neither had had previously detected antigenaemia). The geometric mean CF titre (GMT) for the mothers of the seven babies with either transient antigenaemia or antibody was 1 : 5.9 $\pm$ 3.1 ($\pm$ 1 s.d.), whereas the GMT for those who became carriers was 1 : 46.1 $\pm$ 2.8 ($\pm$ 1 s.d.). Only seven siblings (9.5%) had anti-HBs-positive sera (one of 27 with a carrier father and six of 47 with an antibody-positive father).

Blumberg's autosomal recessive genetic hypothesis for predisposition to the HBsAg carrier state has evoked considerable controversy⁸. The data from several large population studies are compatible with such an explanation for the observed carrier segregation^{1,9–11}. Mazzur has reported retrospective evidence of vertical transmission among Blumberg's original populations¹². These data, in spite of the evidence of vertical transmission, remain compatible with an autosomal recessive mode of inheritance for carrier predisposition, although again lacking two carrier parent families.

Such family clustering of HBsAg carriers could also be explained, as Lederberg and Petrakis pointed out, by the spread of an infectious agent through close personal contact, without any genetic influence^{13,14}. Information from families with the critical matings of two HBsAg-carrier parents has been difficult to obtain, especially in countries with a low carrier prevalence. During a study of vertical transmission of HBV from antigen-carrier mothers to their newborn offspring in Taiwan (a high prevalence area), we found a large number of families with two antigen carrier parents. Data from these families, compared with those with a carrier mother and an antigen-negative, antibody-positive father, were incompatible with an autosomal recessive mode of inheritance. Whether the father was an HBsAg carrier or was antibody positive had no influence on the development of chronic antigenaemia in the newborn children or its existence in their siblings. The maternal CF titre, on the other hand, did influence the baby's response to HBV infection (transient antigenaemia, the development of anti-HBs, or a chronic HBsAg carrier state). In addition, we have reported previously that virtually 100% of babies and their siblings were chronic HBsAg carriers when the mother's serum had a very high CF titre ($\geq$ 1 : 64) (ref. 2).

Our findings suggest that all these children could become

Table 1 Hepatitis B antigenaemia among offspring of HBsAg carrier mothers: influence of HBsAg or anti-HBs status of fathers

Father	Babies		Siblings	
	No. of HBsAg positive/ total*	% Positive	No. of HBsAg positive/ total*	% Positive
HBsAg positive	13/22	59.1	17/27	63.0
Anti-HBs positive	30/53	56.6	27/47	57.4

* $P > 0.5$ by χ^2 .

HBsAg carriers, regardless of their genetic background. Development of the carrier state was probably influenced more by the dose of infectious virus they received (as suggested by the relationship between antigenaemia among the children and their mothers' CF titres). Schweitzer made a similar observation in his finding that vertical transmission from HBsAg carrier mothers in the USA only occurred when the mother's serum contained the e antigen (thought to be related to HBV infectivity)^{15,16}. The early age of exposure in these children may also be important, since the age when infection occurs has been shown to influence the development of the carrier state in other settings¹⁷. Our data support this concept, since very few children had anti-HBs, indicating that when HBV infection occurred early in life, nearly all who were infected became carriers.

In summary, the data presented here seem to be incompatible with an autosomal recessive mode of inheritance for HBsAg carrier predisposition. It seems instead that the infectious dose and the age when infection occurs are more important determinants of the carrier state. The discrepancy between our findings and those of population studies may also indicate that family segregation analyses, especially of a contagious disease, cannot readily discriminate between the roles of environmental and genetic factors¹⁸.

This study was supported through funds provided in part by the Naval Medical Research and Development Command, US Navy Department, and the Office of Naval Research.

CLADD E. STEVENS*
R. PALMER BEASLEY†

Department of Epidemiology and Clinical Investigation,
US Naval Medical Research Unit No. 2,
Taipei, Taiwan and
School of Public Health and Community Medicine,
University of Washington,
Seattle, Washington 98195

Received December 8, 1975; accepted February 12, 1976.

* Present address: Laboratory of Epidemiology, New York Blood Center, 310 East 67th Street, New York, New York 10021.

† To whom reprint requests should be addressed at NAMRU-2, Box 14, APO SF 96263.

- Blumberg, B. S., Friedlaender, J. S., Woodside, A., Sutnick, A. I., and London, W. T., *Proc. natn. Acad. Sci. U.S.A.*, **62**, 1108-1115 (1969).
- Stevens, C. E., Beasley, R. P., Tsui, J., and Lee, W. C., *New Engl. J. Med.*, **292**, 771-774 (1975).
- Beasley, R. P., Stevens, C. E., Shiao, I. S., and Meng, H. C., *Lancet*, **ii**, 740-741 (1975).
- Walsh, J. H., Yalow, R., and Berson, S. A., *J. infect. Dis.*, **121**, 550-554 (1970).
- Schmidt, N. J., Lennette, E. H., in (edit. by Horsfall, F. L., and Tamm, I.) *Viral and Rickettsial Infections of Man*, 1217-1222 (Lippincott, Philadelphia, 1965).
- Shulman, N. R., and Barker, L. F., *Science*, **165**, 304-306 (1969).
- Vyas, G. N., and Shulman, N. R., *Science*, **170**, 332-333 (1970).
- Vyas, G. N., *Nature*, **248**, 159-160 (1974).
- Szmuness, W., Harley, E. J., and Prince, A. M., *Am. J. med. Sci.*, **270**, 293-304 (1975).
- Cepellini, R., Bedarida, G., Carbonara, A. O., Tranchieri, G., and Filippi, G., *Atti Convegni Farmitalia: Antigene Australia ed Epatite Virale*, 53-54 (Minerva Medica, Torino, 1970).
- Grossman, R. A., et al., *Am. J. Epidemiol.*, **101**, 144-159 (1975).
- Mazur, S., Blumberg, B. S., and Friedlaender, J. S., *Nature*, **247**, 41-43 (1974).
- Lederberg, J., in *Hepatitis and Blood Transfusion* (edit. by Vyas, G. N., Perkins, H. A., and Schmidt, R.), 89-93 (Grune and Stratton, New York, 1971).
- Petrakis, N., in *Hepatitis and Blood Transfusion* (edit. by Vyas, G. N., Perkins, H. A., and Schmidt, R.), 85-87 (Grune and Stratton, New York, 1971).
- Schweitzer, I. L., Edwards, V. M., and Brezina, M., *New Engl. J. Med.*, **293**, 940 (1975).
- Magnius, L. O., and Espmark, J. A., *J. Immun.*, **109**, 1017-1021 (1972).
- Szmuness, W., and Prince, A. M., *Am. J. Epidemiol.*, **94**, 585-595 (1971).
- Lillienfeld, A. M., *Cancer Res.*, **25**, 1330-1335 (1969).

Human foetal pituitary peptides and parturition

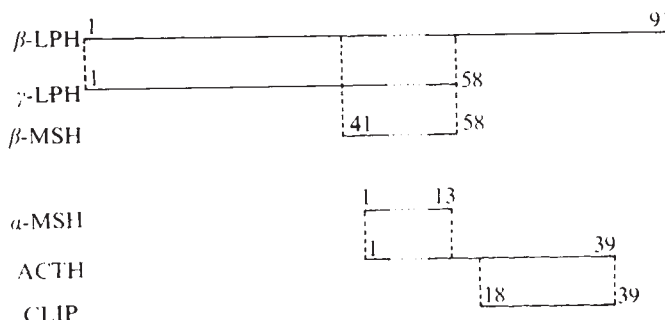
THE pituitary gland produces a family of related peptide hormones (Fig. 1). Adrenocorticotrophic hormone (ACTH), β -lipotropic hormone (β -LPH) and possibly γ -lipotropic hormone (γ -LPH) are produced by the pars distalis. α -Melanocyte stimulating hormone (α -MSH), β -melanocyte stimulating hormone (β -MSH) and corticotropin-like intermediate lobe peptide (CLIP)¹, however, are thought to be formed in the pars intermedia², their being found only in those species in which this structure can be distinguished. They have not been identified in man³. The human foetus, however, does develop a rudimentary pars intermedia which involutes shortly after birth⁴. We present, for the first time, evidence for the occurrence of peptides resembling α -MSH and CLIP in the human foetal pituitary in the second half of pregnancy, and we suggest that these peptides are the dominant hormones of foetal life, only replaced by ACTH before parturition.

Human foetal pituitaries were dissected at autopsy following stillbirth, hysterotomy or neonatal death and adult human pituitaries obtained *post mortem*.

Fractionation of an adult pituitary extract on Sephadex G-100 showed two peaks of large molecular weight material reacting in the β -MSH assay and running in the expected positions of β - and γ -LPH (Fig. 2). There was no significant amount of low molecular weight material corresponding to the position of authentic (1-18) β -MSH. A substantial peak of material reacting in both the C- and N-terminal ACTH assays corresponded to intact ACTH, whereas a shoulder on the trailing edge of this peak contained C-terminal activity which could represent a small amount of CLIP. The final small peak of N-terminal activity can probably be attributed to an α -MSH-like peptide; the melanocyte-stimulating potency of this peak was greatly in excess of that in the ACTH peak but not as large as could be expected for authentic α -MSH.

Fractionation of a pituitary extract from a 24-week female foetus from whom the pituitary was dissected within one hour

Fig. 1 Corticotropin, lipotropin and melanocyte stimulating hormones. β -LPH is a 91-amino acid peptide whose first 58 amino acids are identical in sequence to γ -LPH. β -MSH shares its 18-amino acid sequence with both β - and γ -LPH. ACTH is a 39-amino acid peptide whose first 13 amino acids are identical in sequence with α -MSH, and whose 18-39 sequence is identical to CLIP. . . . The common heptapeptide core responsible for melanocyte stimulating activity in β -LPH, γ -LPH, β -MSH, α -MSH and ACTH.



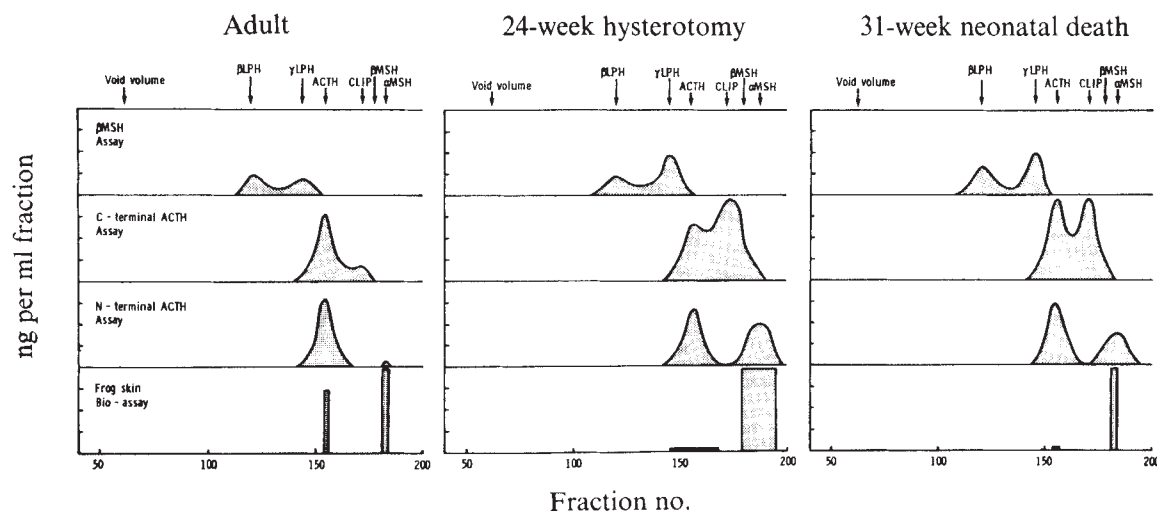


Fig. 2 Chromatography of pituitary extracts on Sephadex G-100. Pituitary fractionations from an adult, a 24-week foetus delivered by hysterotomy and a 31-week infant who survived 12 h post partum. Pituitaries were homogenised in 0.02 M HCl containing 10^{-3} M diisopropylphosphorofluoridate. After centrifugation a sample from the supernatant was applied to a column (1.6×100 cm) of Sephadex G-100 and eluted in 1 M acetic acid. The fractions were analysed in three systems of radioimmunoassay: a β -MSH assay, a C-terminal ACTH assay and an N-terminal ACTH assay. The β -MSH assay⁵ reacted with β -MSH, β -LPH and γ -LPH. The C-terminal ACTH assay⁶ reacted with ACTH and CLIP. The N-terminal ACTH assay⁷ reacted with ACTH and α -MSH. Selected fractions encompassing the peaks for α -MSH and ACTH were also examined for melanocyte stimulating activity by an adaptation of the frog skin bioassay⁸. In this assay the relative potency of α -MSH to ACTH is 1,000:3. Arrows indicate expected positions of β -LPH, γ -LPH, ACTH, CLIP, β -MSH and α -MSH for each of the chromatograms.

of hysterotomy also showed large molecular weight material reacting in the β -MSH assay representing β - and γ -LPH (Fig. 2). As with the adult there was no evidence of low molecular weight authentic β -MSH. A peak of material reacting in both the C- and N-terminal ACTH assays corresponded to intact ACTH. This, however, was followed by a more substantial peak of material which contained solely C-terminal activity and a significant peak of low molecular weight N-terminal activity. The biological potency of the latter was again far in excess of that which could be accounted for by ACTH.

A pituitary extract from a 31-week male infant who died 12 h after a normal delivery was fractionated and demonstrated a pattern intermediate between that of the adult and that of the 24-week foetus (Fig. 2). The peak of material in the C- and N-terminal ACTH assays which corresponded to intact ACTH was approximately equal to the peak of material with solely C-terminal activity and exceeded the peak of low molecular weight N-terminal activity. This last peak again displayed high biological potency. There were peaks representing β - and γ -LPH as before, but no suggestion of any activity that could represent β -MSH.

Extracts of a further four foetal pituitaries were analysed on a BioGel P6 column which gives a clear distinction between ACTH and CLIP. Two stillborn foetuses at 33 and 38 weeks showed a striking preponderance of C-terminal activity over intact ACTH (Fig. 3). In this they resembled the 24-week foetus delivered by hysterotomy (Fig. 2). The proportion of intact ACTH, however, was much increased in a full-term stillborn foetus and was the dominant hormone in a full-term infant who survived 5 h. In two adult pituitaries (69 and 71 yr) there was a large preponderance of ACTH over material resembling CLIP.

Since the 31-week premature infant who survived 12 h (Fig. 2) resembles the two full-term foetuses rather than those of 33 and 38 weeks, it may be that the changing ratio of the small C-terminal peptide to intact ACTH seen in Fig. 3 is associated with parturition rather than with advancing gestational age.

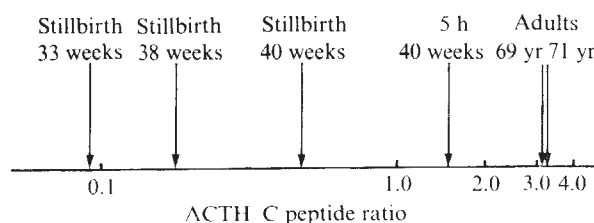
These studies demonstrate that for the most part of intra-uterine life the human foetal pituitary contains only small amounts of intact ACTH, but relatively larger quantities of peptides closely resembling α -MSH and CLIP. It seems probable that the presence of these small molecules reflects a physiological function in the foetus which is not present in postnatal life. It is

of interest to consider this finding in the light of other information on the development of pituitary-adrenal function in the human foetus.

The foetal zone of the human adrenal cortex achieves great size during gestation. At parturition it undergoes rapid involution and the definitive zone hypertrophies. It is commonly argued^{9,10} that foetal pituitary ACTH is the trophic hormone for both the foetal and definitive zones of the adrenal cortex. ACTH administered exogenously, however, is not itself capable of stimulating the foetal zone¹¹, and our findings suggest that ACTH is not produced in large quantities by the human foetal pituitary during the major part of gestation. That the trophic hormone responsible for the growth of the foetal zone might be one of the small peptides identified here, for the first time, is an attractive hypothesis. We have demonstrated that near parturition there is a sharp increase of intact ACTH relative to the smaller peptides. If this were reflected in plasma concentrations the metamorphosis of the foetal adrenal gland during this period would be explicable. The transference of function from foetal to definitive zone would accompany the switch in peptide synthesis by the pituitary.

It is the physiological significance of these events, however, which is of most interest. Liggins has demonstrated that the key event preceding parturition in sheep is a surge of foetal cortisol from the foetal adrenal cortex¹². This event not only helps prepare the foetal lamb for extrauterine existence (for example the production of surfactant in the foetal lung) but it is itself the

Fig. 3 Ratio of ACTH activity to small C-terminal peptide activity in adult and foetal pituitaries as assessed by chromatography on BioGel P6. The foetal pituitaries were obtained from a 33-week stillbirth, a 38-week stillbirth, a 40-week stillbirth and a full-term infant who survived 5 h. Adult pituitaries were obtained *post mortem* from patients aged 69 and 71 yr.



trigger for the initiation of parturition. Premature delivery can in fact be induced at any time in the second half of pregnancy by infusing the foetal lamb with ACTH or cortisol.

It is impossible to determine by direct experiment whether the production of cortisol is of similar significance in the human foetus. Cord blood values for cortisol are, however, greater in babies delivered spontaneously than in those delivered by induction or Caesarean section¹³. It has also been demonstrated that the levels of cortisol in amniotic fluid remain low until week 34 of gestation, but then increase sharply until term¹⁴.

The foetal zone of the human adrenal cortex, though highly active in steroidogenesis, is unable to synthesise glucocorticoids. It is therefore probable that the increased cortisol production derives from the definitive zone. If our hypothesis proves correct that one or other of the small peptides resembling α -MSH and CLIP are the trophic hormones for the foetal zone, while intact ACTH drives the definitive cortex, then the increase of ACTH relative to the small peptides which we have described near parturition would seem to be functionally related to the increasing production of cortisol by the definitive zone of the foetal adrenal gland. Thus the switch in peptide synthesis may be the central mechanism in the chain of events that leads from the passive placenta-dependent foetus to the viable responsive newborn baby.

We thank Drs R. F. Alexander, M. S. Ross, Morag Ellison, Sylvia Lawler, Professor P. Huntingford, Mr A. J. Woolf and the staff of the tissue bank of the Royal Marsden Hospital for the provision of foetal pituitaries, and Dr A. P. Scott for most valuable advice regarding the extraction procedure. R.E.S. was supported by the Wellcome Trust and is at present an MRC Research Fellow.

R. E. SILMAN*
T. CHARD

Department of Reproductive Physiology

P. J. LOWRY

Department of Chemical Pathology,
St Bartholomew's Hospital Medical College,
London EC1A 7BE, UK

IVOR SMITH
I. M. YOUNG

Courtauld Institute,
Middlesex Hospital Medical School,
London, UK

Received December 31, 1975; accepted February 27, 1976.

* To whom reprint requests should be addressed.

- ¹ Scott, A. P., Lowry, P. J., Bennett, H. P. J., McMartin, C., and Ratcliffe, J. G. *J. Endocr.*, **61**, 369–380 (1974).
- ² Lowry, P. J., and Scott, A. P., *Gen. comp. Endocr.*, **26**, 16–23 (1975).
- ³ Scott, A. P., and Lowry, P. J., *Biochem. J.*, **139**, 593–602 (1974).
- ⁴ Wingstrand, K. G., in *The Pituitary Gland*, 3 (edit. by Harris, G. W. and Donovan, B. T.), 1–23 (Butterworths, London, 1966).
- ⁵ Gilkes, J. J. H., et al., *J. clin. Endocr. Metab.*, **40**, 450–457 (1975).
- ⁶ Ratcliffe, J. G., Knight, R. A., Besser, G. M., Landon, J., and Stansfeld, A. G., *Clin. Endocr.*, **1**, 27–44 (1972).
- ⁷ Orth, D. N., in *Methods of Hormone Radio-Immunoassay* (edit. by Jaffe, B. M., and Behrman, H. R.), 125 (Academic, New York and London, 1974).
- ⁸ Shizume, K., Lerner, A. B., and Fitzpatrick, T. B., *Endocrinology*, **54**, 553–560 (1954).
- ⁹ Lanman, J. T., *Pediatrics*, **27**, 140–158 (1961).
- ¹⁰ Eberlein, W. R., in *The Human Adrenal Cortex* (edit. by Christy, N. P.), 317–327 (Harper and Row, New York, 1971).
- ¹¹ Honnebler, W. J., Jobis, A. C., and Swaab, D. F., *J. Obstet. Gynaec. Br. Commonw.*, **81**, 423–438 (1974).
- ¹² Liggins, G. C., Fairclough, R. J., Grieves, S. A., Kendall, J. Z., and Knox, B. S., *Rec. Progr. Hormone Res.*, **29**, 111–159 (1973).
- ¹³ Murphy, B. E. P., *Am. J. Obstet. Gynec.*, **115**, 521–525 (1973).
- ¹⁴ Fencl, M. de M., and Tulchinsky, D., *New Engl. J. Med.*, **292**, 133–136 (1975).

Absolute identification of melatonin in human plasma and cerebrospinal fluid

MELATONIN, the compound responsible for the lightening of amphibian skin, was extracted from bovine pineal glands and chemically characterised by Lerner *et al.*^{1,2}. Recent

reviews^{3–5} have suggested that melatonin has an endocrine role, particularly in relation to chronobiology, pituitary function and cerebral physiology. Most of this work has been accomplished by animal experimentation. So far the evidence for the existence of melatonin in man has been indirect. It is formed biosynthetically from tryptophan and the necessary enzyme system is known to be present in the human pineal throughout life⁷. There is however little evidence to suggest that it circulates in human blood or cerebrospinal fluid (CSF). Attempts to identify melatonin by bioassay⁸ or radioimmunoassay⁹ have not been accompanied by the chemical identification of the substance measured. We present here, for the first time, unequivocal evidence based on gas chromatography (GC) and mass spectrometry (MS) of the presence of melatonin in human plasma and CSF.

Plasmas from normal volunteers and CSF from patients undergoing lumbar puncture as part of other clinical investigations were used. Individual samples were extracted with chloroform. Most lipid and pigment was removed from the chloroform extract by silica gel column or thin-layer chromatography with hexane-ether (80:20) as solvent. Melatonin was eluted with methanol and blown to dryness under argon. Using tritiated melatonin, recoveries in excess of 80% were found at this stage.

Synthetic melatonin was treated with a specific silanising reagent (Sylan BTZ, Supelco) to form the indole *N*-trimethylsilyl derivative. When a Varian Mat 311A mass spectrometer interfaced to a Varian Aerograph 1400 GC was used, the mass spectrum of this derivative gave rise to two major fragments at mass to charge ratio (*m/e*) 232 and 245 and a molecular ion at *m/e* 304 of lower intensity (Fig. 1).

Plasma and CSF extracts were silanised similarly. Pooled plasma extracts (3-ml equivalents) yielded a component with GC retention time and mass spectrum identical with those of synthetic melatonin (Fig. 1). Individual plasma and CSF extracts equivalent to 0.2 ml were subjected to high resolution mass fragmentography at *m/e* 232 and 245 to give, at the correct GC retention time for melatonin, peaks with accurate mass identical to those of synthetic melatonin.

The elucidation of any physiological or even pathological role for melatonin in man requires first its identification and subsequently its quantification. Our results constitute the first absolute identification of melatonin in human plasma and CSF. Moreover, the major ions 232 and 245, since they carry a large proportion of the total ion current of silanised melatonin, may be utilised in a single ion detection procedure for sensitive quantitative analysis. Thus GCMS in demonstrating the existence of circulating melatonin in man provides at the same time a method for its assay.

P.E.M. and R.E.S. are supported by MRC research fellowships.

IVOR SMITH

Courtauld Institute,
Middlesex Hospital Medical School,
London W1P 5PR, UK

P. E. MULLEN

Maudsley Hospital and Institute of Psychiatry,
London SE5, UK

R. E. SILMAN

Department of Reproductive Physiology,

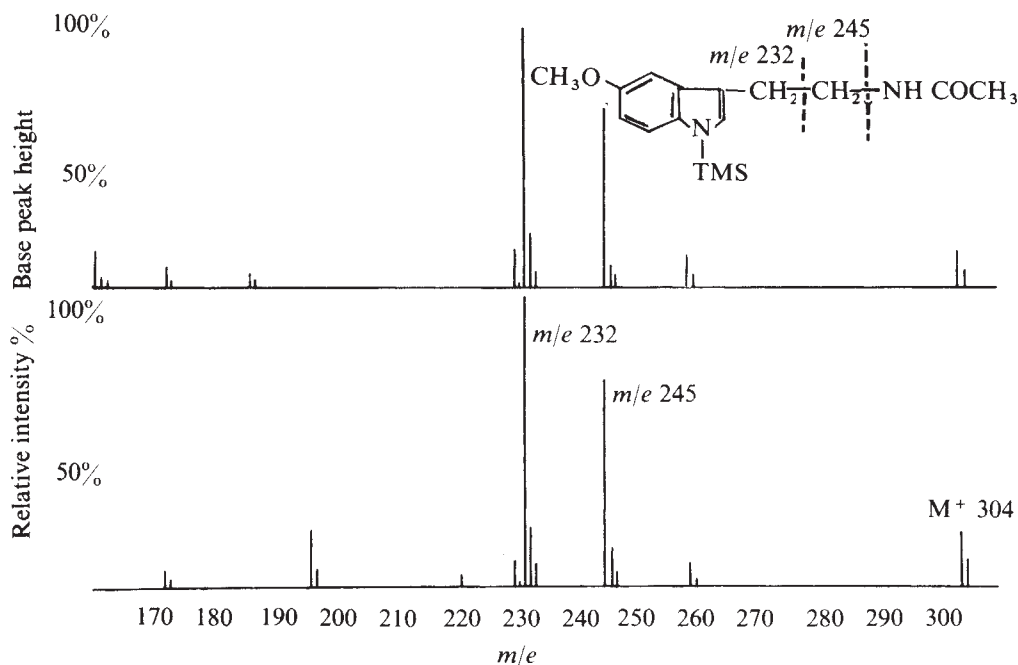
W. SNEDDEN

B. W. WILSON

Department of Chemical Pathology,
Mass Spectrometry Group,
St Bartholomew's Hospital Medical College,
London EC1, UK

Received February 24; accepted March 11, 1976.

Fig. 1 Partial mass spectra obtained from silanised synthetic melatonin and silanised plasma extract. The bonds which fragment to give rise to the two major ions are indicated.



- ¹ Lerner, A. B., Case, J. D., Takahashi, Y., Lee, T. H., and Mori, W., *J. Am. chem. Soc.*, **80**, 2587 (1958).
- ² Lerner, A. B., Case, J. D., and Heinzelman, R. V., *J. Am. chem. Soc.*, **81**, 6085 (1959).
- ³ Cardinali, D. P., in *Current Topics in Experimental Endocrinology*, 2 (edit. by James, V. H. T., and Martini, L.), 107-128 (Academic, New York, London, 1974).
- ⁴ *The Pineal Gland* (edit. by Wolstenholme, G. E. W., and Knight, J.) (Churchill Livingstone, London, 1971).
- ⁵ Axelrod, J., *Science*, **184**, 1341-1348 (1974).
- ⁶ Reiter, R. J., in *MTP International Review of Science, Physiology, Series 1, 5, Endocrine Physiology* (edit. by McCann, S.), 277-305 (Butterworths, London, 1974).
- ⁷ Wurtman, R. J., Axelrod, J., and Barchas, J. D., *J. clin. Endocr. Metab.*, **24**, 299-301 (1964).
- ⁸ Pellham, R. W., Vaughan, G. M., Sandock, K. L., and Vaughan, M. K., *J. clin. Endocr. Metab.*, **37**, 341-344 (1973).
- ⁹ Arendt, J., Paunier, L., and Sizonenko, P., *J. clin. Endocr. Metab.*, **40**, 347-349 (1975).

Protection of spermatogenesis by α_{2u} globulin in rats treated with oestrogen

α_{2u} GLOBULIN is an androgen-dependent rat urinary protein with a molecular weight of about 23,800 (ref. 1). In the male rat, starting from puberty until senescence, α_{2u} globulin is synthesised and secreted by the hepatic tissue and is rapidly cleared by the kidneys into the urine²⁻⁴. In addition to its urinary concentration α_{2u} globulin has been found to be concentrated by the salivary glands and released into the saliva (A.K.R., and J.G.B., in preparation). Synthesis of this protein could be induced in the spayed female rat by androgen, and oestrogenic compounds completely inhibited α_{2u} synthesis in the mature male rat⁵. But α_{2u} globulin does not bind testosterone, oestradiol-17 β or any of their metabolites (unpublished results of A.K.R.). Pretreatment of mature male rats with oestrogen induces a dose-dependent temporary lag period within which α_{2u} globulin is not synthesised even under androgenic stimulation. With a daily oestradiol dose of 50 μ g per 100 g body weight a lag period extending up to 3-4 weeks has been established⁶. Although the adult male rat synthesises 20-30 mg of α_{2u} globulin per day, no physiological role of this protein has yet been established. The close parallel between the male reproductive capability and the production of α_{2u} globulin^{4,5} prompted us to look into the possible reproductive role of this protein. The discovery of the oestrogen-mediated lag period in the mature male rats also provided an opportunity

to look into the spermatogenic abnormalities within this lag period and any reparative effects which α_{2u} globulin may have on such damage. The results reported here demonstrate considerable protection of spermatogenesis by α_{2u} globulin against potential damage by oestrogen.

α_{2u} Globulin was prepared from male rat urine as before⁷. The protein was purified further on a Sephadex G-75 column. α_{2u} Globulin prepared according to the above procedure gave a single electrophoretic band on sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis. Because investigations by Lacy and Lofts⁸ have indicated that damage to spermatogenesis caused by oestrogen could be reversed by follicle stimulating hormone (FSH), it was considered important to rule out any possible trace of contamination of FSH in our α_{2u} preparation.

Potent antiserum against FSH-B was prepared by repeated injection of the hormone with Freund's complete adjuvant into rabbits, using a procedure outlined for preparation of antiserum against α_{2u} globulin⁷. The antiserum prepared according to the above procedure, when tested on Goodman plates⁹, produced four precipitin lines against 25 μ g of FSH-B and three precipitin lines (one major and two minor) with the same amount of FSH-1. Even 100-fold excess of α_{2u} globulin (2.5 mg) showed no precipitin reaction against the antiserum. Double antibody radioimmunoassay¹⁰ of a solution (2.5 mg ml⁻¹) of α_{2u} globulin preparation under various dilutions with the anti-FSH and FSH-1 system also failed to detect any trace contamination of FSH in the α_{2u} preparation. Antiserum prepared against α_{2u} globulin also yielded no precipitin reaction when tested against whole pituitary homogenate.

Mature male rats of Yale strain (about 100 d old at the start of the experiments) were used for this investigation. Thirty-five animals were treated daily with oestradiol-17 β for 7 d (50 μ g per 100 g body weight per d). At the end of treatment, five animals were killed and their testes were fixed in Bouin's fixative and used for paraffin embedding. The remaining 30 animals were divided into two groups, one of which received injections of 0.2 ml of 0.14 M NaCl per day and the other received subcutaneous injections of α_{2u} globulin (1 mg in 0.2 ml of 0.14 M NaCl per animal per day) for the rest of the experimental period. Five animals from each group were killed at 7-d intervals and their testicular tissue was fixed in Bouin's fixative and used for

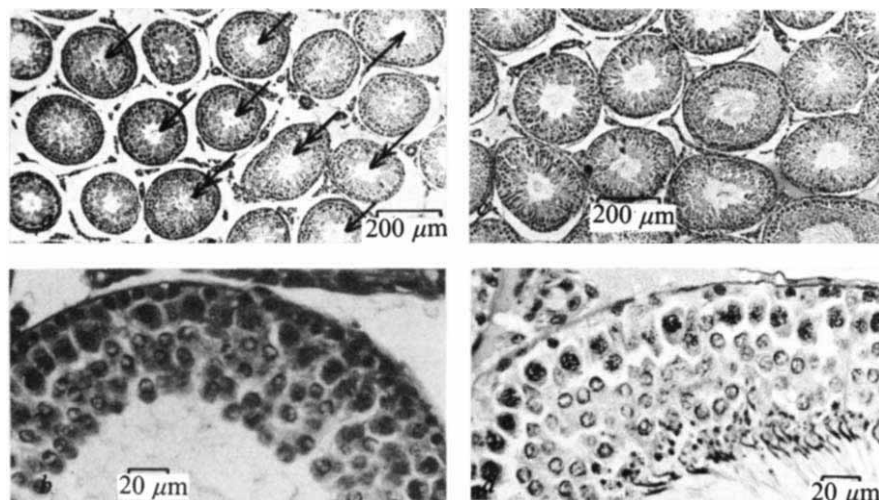


Fig. 1 Photomicrographs of PAS-haematoxylin-stained sections of the testes from animals treated with oestrogen alone and with oestrogen followed by α_{2u} globulin. *a*, Testicular sections from control animals receiving daily injections of oestradiol-17 β for the first 7 d followed by injections of 0.2 ml of 0.14 M NaCl (placebo) per day for the next 3 weeks. Tubules marked with $\rightarrow$ show absence of step 15 to step 19 spermatids. Tubules marked with $\rightarrow\rightarrow$ show drastic reduction of step 9 to step 14 spermatids. Considerable shrinkage of the tubular diameter could also be noted when compared with (c). *b*, Treatment same as for (a). This photomicrograph shows the cross section of a tubule at stage VIII of the seminiferous cycle. Note the absence of step 19 spermatids. *c*, Testicular section from an animal treated daily with oestradiol-17 β for the first 7 d, followed by daily treatment with α_{2u} globulin for the next 3 weeks. The tubular sections show normal spermatogenesis with normal cellular associations at various stages of the cycle. Unlike the section shown in (a), no tubular shrinkage could be seen. *d*, Treatments same as for (c). This photomicrograph shows the cross section of a tubule at stage VIII of the seminiferous cycle. In contrast to (b) the tubule shows normal spermatogenesis proceeding up to mature step 19 spermatid.

paraffin embedding. Sections (5 μ m thick) of the testes were stained with periodic acid-Schiff (PAS)-haematoxylin and used for histological evaluation^{11,12}.

The stained sections from animals treated for 7 d with oestradiol-17 β and killed on the eighth day showed no discernible change in spermatogenesis. Testicular sections from animals treated for 7 d with oestradiol and another week of rest period (no hormone treatment) also showed almost normal morphology. Tubular shrinkage and degenerative changes were found to appear at the second week after the last oestradiol treatment. Seminiferous epithelium in these groups of animals (2 weeks after last oestradiol treatment)

Table 1 Quantitative evaluation of the oestrogenic damage and its prevention by α_{2u} globulin of the seminiferous epithelium at various stages of the spermatogenic cycle

Cellular associations (stages of the cycle)	Index of degeneration in group 1 (oestrogen alone)			Index of degeneration in group 2 (oestrogen followed by α_{2u} globulin)		
	Day 14	Day 21	Day 28	Day 14	Day 21	Day 28
I	0	80	72	0	0	0
II	0	25	58	0	0	0
III	0	62	73	0	0	0
IV	0	66	100	0	0	0
V	0	55	90	0	0	0
VI	6	42	93	0	0	0
VII	7	82	100	0	0	0
VIII	0	50	100	0	0	0
IX	8	83	19	0	0	0
X	0	83	11	0	0	0
XI	0	54	16	0	0	0
XII	4	72	42	0	0	0
XIII	0	72	42	0	0	0
XIV	0	40	31	0	0	0

All animals were treated for the first 7 d with oestradiol-17 β (50 μ g in 0.1 ml vehicle per 100 g body weight per day; subcutaneous injection of the hormone as an emulsion in 0.1 M sodium phosphate, pH 7.2, propylene glycol and Tween 80, 89.6:10.0:0.4, by volume). After initial oestrogen treatment, the animals in group 1 received no other treatment except daily injections of 0.2 ml of saline (0.14 M NaCl). Beginning from the 8th day (1 d after the last oestrogen injection) until death, the animals in group 2 received daily injections of α_{2u} globulin (1 mg per rat per d, subcutaneously in 0.14 M NaCl). The tubular cross sections showing the presence of degenerating germinal cells and/or containing less than 50% of any particular cell type within the specific cellular association were considered to be abnormal. Percentage tubular cross section showing abnormality is used as the index of degeneration.

$$\text{Index of degeneration} = \frac{\text{No. of abnormal tubular cross sections}}{\text{Total no. of tubular cross sections examined}} \times 100$$

The fractional numbers are brought to the nearest whole number.

showed drastic reduction and in some sections complete absence of spermatids of step 9 to step 19. Some of the pachytene spermatocytes within stages VIII through XIII also showed degenerative changes. The testicular sections from animals killed 3 weeks after the last oestradiol treatment showed relatively small and atrophic tubules. Qualitatively, spermatogenesis was found to progress only up to step 7-8 spermatids. Step 9-19 spermatids were found to be absent in most of the tubules and in some instances when present were invariably abnormal (Fig. 1a and b). Quantitative evaluation of the abnormality within the various stages of the seminiferous cycle are summarised in Table 1.

The group of animals given daily injections of α_{2u} globulin after initial treatment with oestradiol-17 β showed no degenerative changes in their testicular sections throughout the period of investigation. The testicular sections from this group of animals 1, 2 and 3 weeks after the last oestradiol treatment showed an intact seminiferous cycle leading to normal spermatogenesis (Fig. 1c and d; Table 1).

In another series of experiments, adult male rats after the first 7 d of oestrogen treatment were subsequently treated with testosterone alone (250 μ g per 100 g body weight) and testosterone plus α_{2u} globulin. Although testosterone, together with α_{2u} globulin, could provide complete protection against the oestrogen-induced damage to spermatogenesis, testosterone alone failed to do so.

The results presented here show that α_{2u} globulin can provide protection against oestrogen-mediated damage of spermatogenesis in the rat. We have shown before that oestrogen treatment completely stops the synthesis of α_{2u} globulin within 7 d, and with a dose of 50 μ g per 100 g body weight reinitiation of α_{2u} synthesis does not start until 3-4 weeks later⁶. Therefore, any possible spermatotropic role of α_{2u} globulin is expected to be missing within the oestrogen-mediated lag period. The observed damage to spermatogenesis, in particular the failure of spermatid maturation in the oestrogen-treated group, and the prevention of such inhibitory effect by α_{2u} globulin within the lag period, may indicate a physiological role of α_{2u} globulin in spermatogenesis.

We thank Dr V. N. Reddy, Dr N. J. Unakar and Marian Wilson for criticisms and suggestions. Rat FSH and anti-FSH preparations were given by the NIAMDD rat pituitary hormone distribution program of the National Institute of

Arthritis, Metabolism and Digestive Diseases. This work was supported by a grant from the National Institutes of Health.

ARUN K. ROY*
JENNIE G. BYRD

Department of Biological Sciences,
Oakland University,
Rochester, Michigan 48063

NAREN M. BISWAS

Department of Physiology,
University of Calcutta,
Calcutta-9, India

AJIT K. CHOWDHURY

Department of Reproductive Biology
and Endocrinology,
University of Texas Medical School,
Houston, Texas 77025

Received December 15, 1975; accepted February 19, 1976.

*To whom correspondence should be addressed.

- 1 Roy, A. K., Neuhaus, O. W., and Harmison, C. R., *Biochim. biophys. Acta*, **127**, 72-81 (1966).
- 2 Roy, A. K., and Neuhaus, O. W., *Biochim. biophys. Acta*, **127**, 82-87 (1966).
- 3 Roy, A. K., and Raber, D. L., *J. Histochem. Cytochem.*, **20**, 89-96 (1972).
- 4 Roy, A. K., Milin, B. S., and McMinn, D. M., *Biochim. biophys. Acta*, **354**, 213-232 (1974).
- 5 Roy, A. K., and Neuhaus, O. W., *Nature*, **214**, 618-620 (1967).
- 6 Roy, A. K., McMinn, D. M., and Biswas, N. M., *Endocrinology*, **97**, 1501-1508 (1975).
- 7 Roy, A. K., *J. Endocr.*, **56**, 295-301 (1973).
- 8 Lacy, D., and Lofts, B., *Proc. R. Soc.*, **162**, 188-197 (1965).
- 9 Goodman, M., *Hum. Biol.*, **34**, 104-150 (1962).
- 10 Midgley, A. R., Jr., *J. clin. Endocr. Metabolism*, **27**, 295-299 (1967).
- 11 Leblond, C. P., and Clermont, Y., *Ann. N.Y. Acad. Sci.*, **55**, 548-573 (1952).
- 12 Perey, B., Clermont, Y., and Leblond, C. P., *J. Anat.*, **108**, 47-77 (1961).

Biochemical responses in lungs of ozone-tolerant rats

THE health hazard of ozone, the principal oxidant air pollutant of photochemical smog, has been well documented¹⁻³. One of the notable features of ozone toxicity is the development of tolerance. When animals are exposed to a single sublethal dose of ozone they become resistant to subsequent lethal exposure. This tolerance phenomenon has been demonstrated in several species, and the effect has been shown to persist for well over one month³⁻⁶. One of its toxic effects has been attributed to its oxidative

nature, and, more specifically, the initiation of lipid peroxidation⁷⁻⁹. Glutathione peroxidase (GP), which uses the reducing equivalent of reduced glutathione (GSH) to catalyse the decomposition of toxic hydroperoxides, and metabolically related enzymes glutathione reductase (GR) and glucose-6-phosphate dehydrogenase (G6PD), have been suggested as a key mechanism for protection of cellular components from deleterious effects of hydroperoxides^{10,11}. The activities of these enzymes have been found to increase in the lungs of rats exposed to chronic levels of ozone⁹, and in tissues of rats subjected to other oxidative stress^{12,13}. Thus changes in the activities of these enzymes seem to be related in some way to the defence mechanism against peroxidative damage. In this study, rats tolerant to ozone maintained higher activities of GP, GR and G6PD, and higher levels of GSH in the lungs following excessive exposure. The adaptive response of this potentially protective enzymic system in the lungs of ozone-tolerant rats may therefore be partly responsible for their resistance against lethal exposure.

Seventy-day-old male Sprague-Dawley rats (chronic respiratory disease free) were initially exposed to either 0 or 0.76 p.p.m. ozone for 3 d (to develop tolerance) in monitored chambers as described previously⁹. After 8 d of recovery in an air-filtered room, portions of tolerant and non-tolerant animals were exposed to either 0 or 4.0 p.p.m. ozone for 8 h. Sixteen hours later surviving animals and their respective controls were killed by exsanguination following pentobarbital injection. After perfusing with saline through the pulmonary artery to remove blood, the lungs were homogenised, fractionated and assayed for the activities of GP, GR, G6PD, and levels of protein and GSH as described previously^{9,14}.

Sixteen hours after the end of the acute exposure, five out of ten non-tolerant rats died compared with only one out of seven of the tolerant group. The lungs of both acute exposed groups exhibited severe haemorrhage and oedema, but the tolerant animals showed less severe injury. The biochemical responses of the lungs are summarised in Table 1. Exposure of rats to 4.0 p.p.m. ozone for 8 h resulted in a marked increase in protein content, and a significant ($P < 0.05$) decrease in the activity of GP, GR and G6PD, and levels of GSH in the lungs of both tolerant and non-tolerant groups. Tolerant animals, however, showed significant, higher activities (from Student's *t* test) of GP, GR and G6PD, and higher GSH levels in the lungs than

Table 1 Biochemical responses in the lungs of rats exposed to ozone 16 h after the end of exposure to 4 p.p.m. ozone for 8 h

Biochemical parameter	Tolerance*	Exposed (mean $\pm$ s.d.)	Control	P
Protein (mg per lung)	Yes	136 $\pm$ 6	71 $\pm$ 4	<0.001
	No	152 $\pm$ 17	73 $\pm$ 3	<0.001
		P NS	NS	
Reduced glutathione (nmol non-protein sulphhydryl per mg protein)	Yes	15 $\pm$ 2	31 $\pm$ 2	<0.001
	No	11 $\pm$ 1	29 $\pm$ 4	<0.001
		P <0.01	NS	
Glutathione peroxidase ⁸ (nmol NADPH oxidised or reduced per min per mg protein)	Yes	42 $\pm$ 3	67 $\pm$ 5	<0.001
	No	34 $\pm$ 4	64 $\pm$ 3	<0.001
		P <0.01	NS	
Glutathione reductase ⁸ (nmol NADPH oxidised or reduced per min per mg protein)	Yes	22 $\pm$ 3	40 $\pm$ 5	<0.001
	No	18 $\pm$ 2	39 $\pm$ 2	<0.001
		P <0.02	NS	
Glucose-6-phosphate dehydrogenase ⁸ (nmol NADPH oxidised or reduced per min per mg protein)	Yes	44 $\pm$ 4	60 $\pm$ 5	<0.001
	No	37 $\pm$ 4	59 $\pm$ 3	<0.001
		P <0.02	NS	

*Exposed to 0.76 p.p.m. ozone for 3 d and allowed to recover for 9 d. NS, Not significant ($P > 0.05$).

those of the non-tolerant rats. When the data are expressed as total units per lung, tolerant rats still show a higher level of GSH and higher activities of GP, GR and G6PD. It should be noted that the above biochemical parameters in the lungs of rats recovering for 9 d from the initial exposure were not different from those of the non-exposed controls (Table 1).

In another experiment, 60-d-old male rats were initially exposed to either 0 or 0.73 p.p.m. ozone for 3 d. Eleven days after exposure, 11 tolerant and 13 non-tolerant animals were exposed to 3.9 p.p.m. ozone for 8 h. Only one of the tolerant rats had died 18 h later, as against seven of the non-tolerant group. The lungs of tolerant animals also showed relatively less severe injury and maintained higher activities of GP, GR and G6PD, and higher GSH levels.

Throughout this research, it was observed that the activities of GP, GR and G6PD, and GSH levels in the lungs of animals exposed to 0.8 p.p.m. ozone for 3 d increased significantly above those of the controls, and that these biochemical changes began to revert towards control values in 2 d, reaching control levels about 7 d after recovering in ozone-free air (unpublished observations). Even though the biochemical changes in the lungs of tolerant rats completely returned to control level, significantly higher activities of GP, GR and G6PD, and GSH levels, accompanied by a reduced mortality rate and lung injury were developed when this group of rats were subjected later to the acute exposure. Although the precise significance and specificity of this correlation remain to be seen, it is possible that higher activities of GP, GR and G6PD, and GSH levels in the lungs of tolerant animals may render them more resistant to peroxidation damage induced by ozone. It has been shown that ozone-tolerant animals maintained higher GSH levels in the lungs, and theorised that the activities of pentose shunt pathway may be related to resistance or tolerance to injury from inhaled irritants^{4,13}.

The mechanisms of the differential biochemical response between the tolerant and non-tolerant rats to acute ozone exposure remain to be elucidated. It is possible that the enzymic synthesis machinery in the lungs of tolerant rats remained 'potentially active' even though the enzyme activities reverted to control levels. It has been shown morphologically that ozone primarily affects the type 1 cells of small conducting airways and adjacent alveoli^{16,17}. Following injury to less than 1 p.p.m. ozone, repair processes, as evidenced by the replacement of type 1 cells by the cuboidal epithelial cells in the injured areas, begin within 24 h and reach a maximum between 48 and 72 h. When animals are allowed to recover in room air, the cuboidal cells are gradually transformed into type 1 cells, and the tissue returns to normal in about 7 d (refs 16 and 18). From a morphological point of view, therefore, the newly transformed type 1 cells may be potentially more prone to metabolic change and thus become more resistant to subsequent exposure. Further studies such as morphometric measurement of type 1 and type 2 cells are, however, needed to clarify this view.

The GP system also has an important role in protecting red blood cells against oxidative damage. When red cells genetically lack one or more of these enzymes or enzymes involved in GSH synthesis, they become susceptible to oxidative drugs which induce haemolytic anaemia^{19,20}. The GP system may also be involved in tolerance phenomena induced by other types of oxidative agents, as GP uses a variety of hydroperoxides as substrates¹⁰.

The present observations that tolerant animals are capable of maintaining higher activities of GP, GR and G6PD, and GSH levels in the lungs suggest that this enzymic system may be partly responsible for their resistance against lethal ozone exposure. This study provides evidence for the possible role of the protective GP system in the development of tolerance to oxidative damage.

This work was supported by research grants from the US Public Health Service.

CHING K. CHOW

California Primate Research Center,
University of California,
Davis, California 95616

Received December 22, 1975; accepted February 24, 1976.

- ¹ Dungworth, D. L., Cross, C. E., Gillespie, J. R., and Plopper, C. G., in *Ozone Chemistry and Technology* (edit. by Murphy, J. S. and Orr, J. R.), ch 2, 29-46 (Franklin Institute Press, Philadelphia, 1975).
- ² Menzel, D. B., *A. Rev. Pharmacol.*, **10**, 379-394 (1970).
- ³ Stokinger, H. E., and Coffin, D. L., in *Air Pollution*, **1**, (edit. by Stern, A. C.), 446-546 (Academic, New York, 1968).
- ⁴ Stokinger, H. E., and Scheel, L. D., *Archs Environ. Health*, **4**, 327-334 (1962).
- ⁵ Stokinger, H. E., Wagner, W. D., and Wright, P. G., *Archs Industr. Health*, **14**, 158-162 (1958).
- ⁶ Matzen, R. N., *Am. J. Physiol.*, **190**, 84-88 (1957).
- ⁷ Goldstein, E. B. D., Buckley, R. D., Cardenas, R., and Balchum, O. J., *Science*, **169**, 605-606 (1970).
- ⁸ Roehm, J. H., Hadley, J. G., and Menzel, D. B., *Archs Intern. Med.*, **128**, 88-93 (1971).
- ⁹ Chow, C. K., and Tappel, A. L., *Lipids*, **7**, 518-524 (1972).
- ¹⁰ Little, C., and O'Brien, P. J., *Biochem. biophys. Res. Commun.*, **31**, 145-150 (1968).
- ¹¹ Christophersen, B. O., *Biochim. biophys. Acta*, **164**, 35-46 (1968).
- ¹² Chow, C. K., Reddy, K., and Tappel, A. L., *J. Nutr.*, **103**, 618-624 (1973).
- ¹³ Reddy, K., and Tappel, A. L., *J. Nutr.*, **104**, 1069-1078 (1974).
- ¹⁴ Chow, C. K., Mustafa, M. G., Cross, C. E., and Tarkington, B. K., *Environ. Physiol. Biochem.*, **5**, 142-146 (1975).
- ¹⁵ Fairchild, E. J., *Archs Environ. Health*, **14**, 111-125 (1967).
- ¹⁶ Stephens, R. J., Sloan, M. F., Evans, M. J., and Freeman, G., *Am. J. Pathol.*, **74**, 31-58 (1974).
- ¹⁷ Schwartz, L. W., Dungworth, D. L., Mustafa, M. G., Tarkington, B. K., and Tyler, W. S., *Lab. Invest.*, (in press).
- ¹⁸ Stephens, R. J., Sloan, M. F., Evans, M. J., and Freeman, G., *Expl molec. Pathol.*, **20**, 11-23 (1974).
- ¹⁹ Beutler, E., *Fedn Proc.*, **31**, 141-146 (1972).
- ²⁰ Altman, K. I., *Am. J. Med.*, **27**, 936-951 (1959).

Mutagenic action of ascorbic acid

MICROBIAL¹, mammalian^{2,3} and human^{4,5} bioassays are being applied to uncover the mutagenic action of synthetic compounds which have been introduced accidentally or intentionally into the environment. Somewhat less attention has been given to the potential genetic hazard of chemicals which are an essential part of human nutrition or are formed within mammalian tissues and organs. Ascorbic acid belongs to this group of "neglected" naturally occurring compounds, although its capacity to form radicals and its interaction with viral⁶ and bacterial⁷ DNA should have raised suspicion about its possible mutagenic capacity. Considering the large daily intake of ascorbic acid, its addition to many food products⁸ and its potential use as an inhibitor of the intragastric formation of *N*-nitroso compounds^{8,9} and nitrosamine formation in food products¹⁰, its action on the genome of human cells requires investigation. We here present evidence, from several experimental systems, suggesting that vitamin C, particularly in the presence of copper, has a mutagenic effect.

Several endpoints were used to uncover the effect of freshly prepared ascorbic acid and an ascorbic acid-Cu²⁺ mixture on the genome of cultured human fibroblasts. DNA fragmentation was measured as a shift in sedimentation profile after centrifugation in an alkaline sucrose gradient¹¹, DNA repair was estimated as unscheduled uptake of tritiated thymidine (³H-TdR) into non-dividing cell populations¹² and chromosome aberrations were quantified by counting breaks and exchanges on the first metaphase plate after treatment. Fibroblasts were obtained from small skin biopsies of four healthy 20-26-yr-old females and males and were used at the third to fifth tissue culture passage. The mutagenic capacity of ascorbic acid and an ascorbic acid-Cu²⁺ mixture were tested further using Ames' tester strain TA 100 of *Salmonella typhimurium*¹.

Cultured human fibroblasts treated with a mixture of ascorbic acid (Mallinckrodt) and Cu²⁺ exhibited DNA fragmentation (Fig. 1), DNA repair synthesis (Fig. 2), and chromosome aberrations including chromatid breaks and exchanges (Table 1). An application of CuSO₄ at concentrations varying from 10⁻⁶ to 10⁻⁴ M had no detectable effect

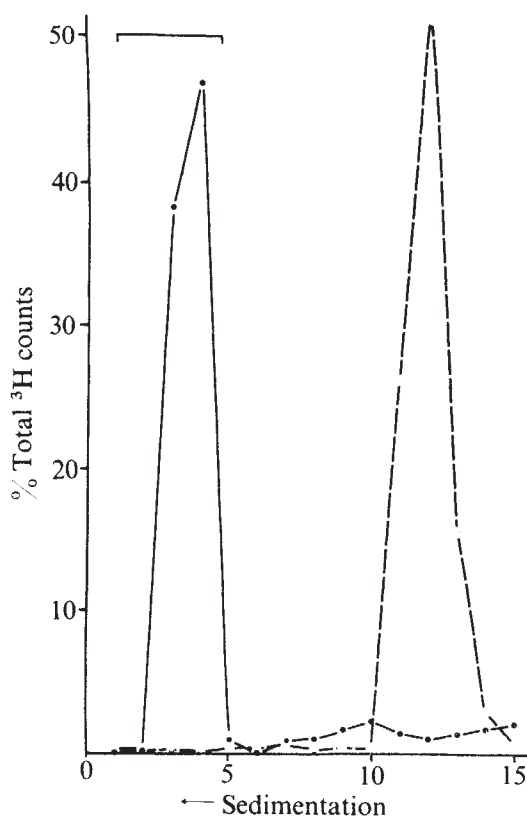


Fig. 1 Alkaline gradient sedimentation profiles of ^3H -DNA released from cultured human fibroblasts. ●, Cells exposed for 2 h to $1.8 \times 10^{-5}\text{M}$ CuSO_4 ; ○, cells exposed for 2 h concurrently to ascorbic acid ($1.8 \times 10^{-3}\text{M}$) and CuSO_4 ($1.8 \times 10^{-5}\text{M}$). Horizontal bar indicates the sedimentation range into which DNA from untreated fibroblasts falls. ^3H -TdR ($1 \mu\text{Ci ml}^{-1}$) was added to the replicating fibroblasts for 20 h; thereafter the cell cultures were exposed to the compounds, rinsed four times with cold 0.024M EDTA/ 0.075M NaCl buffer ($\text{pH } 7.5$), scraped from the Petri dish surface, pelleted (3 min, $1,000\text{ r.p.m.}$), resuspended to give 10^4 – 10^6 cells per $100 \mu\text{l}$, and placed on a lysing solution¹¹ on top of a linear alkaline sucrose gradient (5–20%, 5 ml) which was prepared over a 1-ml cushion of 2.3M sucrose. After a 10-min lysis period at 25°C the gradients were centrifuged at $25,000\text{ r.p.m.}$ for 30 min.

on DNA and chromosomes. If the ascorbic acid solution is flushed with N_2 immediately before addition to cell cultures, the DNA-damaging and chromosome-breaking action is virtually abolished. The active concentrations of ascorbic acid– Cu^{2+} are relatively high as compared with those of the potent carcinogen and mutagen *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) which triggers a DNA repair

Table 1 Metaphase plates of cultured human fibroblasts with chromosome aberrations (breaks and exchanges) after a 2-h exposure to oxidised ascorbic acid and an ascorbic acid CuSO_4 mixture

	Metaphase plates with chromosome aberrations (%) [*]	
	Ascorbic acid + O_2 [†]	Ascorbic acid + Cu^{2+} [‡]
$2 \times 10^{-3}\text{M}$	25.8	— [§]
$1 \times 10^{-3}\text{M}$	28.5	—
$6 \times 10^{-4}\text{M}$	23.4	—
$3 \times 10^{-4}\text{M}$	15.0	26.5
$1.5 \times 10^{-4}\text{M}$	5.0	5.4
Controls	0–15	

Controls were cultured cells untreated with the ascorbic acid preparations.

^{*}Samples were taken 18 h after treatment and 4 h after application of colchicine.

[†] O_2 was flushed through the ascorbic acid solution for 2 h before addition to the cultured fibroblasts.

[‡]The ascorbic acid– CuSO_4 ratio was 1,000:1

[§]Concentrations of $6 \times 10^{-4}\text{M}$ and greater suppressed cell division.

Table 2 Frequency of mutations (*his*⁺ revertants) of *S. typhimurium* (TA 100) after exposure to ascorbic acid, ascorbic acid plus Cu^{2+} or MNNG

	<i>his</i> ⁺ Revertants per 10^6 survivors [*]	
	Ascorbic acid	Ascorbic acid + Cu^{2+}
$3.5 \times 10^{-3}\text{M}$	2.2	160.0
2.5×10^{-2}	1.4	108.7
1.5×10^{-2}	0.7	3.8
1.0×10^{-2}	1.6	1.2
5.0×10^{-3}	1.1	1.0
Control [†]	1.1 (0.5–2.1)	—
		MNNG
		3,090.9
		171.4

^{*}Bacteria in suspension were treated with the various compounds for 15 min at 22°C , washed, centrifuged, resuspended with nutrient broth and plated on histidine-deficient culture medium¹. The ascorbic acid + $\text{CuSO}_4 \times 5\text{H}_2\text{O}$ (1,000:1) solution was incubated for 30 min at 37°C before use.

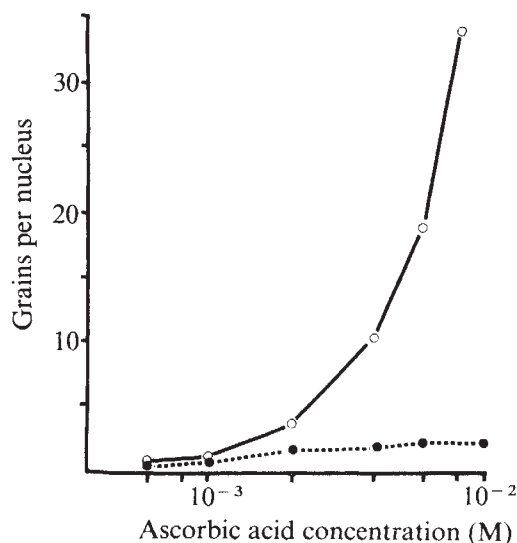
[†]Controls were bacteria untreated with the various chemical preparations. The range of mutation frequencies is shown in parentheses.

synthesis at concentrations varying from 10^{-7} to $2 \times 10^{-4}\text{M}$ (ref. 5), and induces chromosome aberrations at doses of 5×10^{-6} to $6 \times 10^{-5}\text{M}$ (ref. 13), or 4-nitroquinoline-1-oxide (4NQO) which starts to elicit these two phenomena at $2 \times 10^{-7}\text{M}$ and $4 \times 10^{-4}\text{M}$, respectively^{5,14}.

A significant frequency of *his*⁺ revertants per surviving bacterial population was induced by a short (15 min) exposure to the ascorbic acid– Cu^{2+} mixture (Table 2). At equimolar concentrations freshly prepared ascorbic acid does not exert a detectable mutagenic effect. The range of the active concentrations of the ascorbic acid– Cu^{2+} mixture is relatively narrow and thus can be missed easily. For comparison, the mutagenic effect of MNNG at these relatively high concentrations has been added to Table 2. In this connection it should be noted that about 100 new *his*⁺ revertants which require a dose of about $2 \times 10^{-4}\text{M}$ ascorbic acid– Cu^{2+} are induced by as low a concentration of MNNG as 10^{-4}M . Copper ions at the concentrations of 10^{-5} – 10^{-8} that we used had no detectable effect on the mutation frequency of *S. typhimurium*. The latter observation agrees with the lack of mutagenic activity of CuCl and CuCl_2 in the rec assay of *Bacillus subtilis*¹⁵.

Considerable evidence has been accumulated to show that oxidation products of ascorbic acid are mutagenic for

Fig. 2 Unscheduled incorporation of ^3H -TdR into nuclei of non-proliferating cultured human fibroblasts exposed for 90 min to ascorbic acid: CuSO_4 mixtures (○) or an ascorbic acid solution which was bubbled with pure N_2 (Canada Liquid Air Ltd) for 1 h before application to the culture cells (●). The autoradiographic analysis of the incorporated ^3H -TdR and the arrest of proliferation by an arginine-deficient culture medium were described before¹².



microbial and mammalian cells: ascorbic acid in the presence of oxygen converts covalently-closed circular DNA to open circular DNA molecules (unpublished results of A. R. Morgan, R. C. Cone and T. M. Elgert), causes single-strand breaks in DNA⁶, inactivates transforming DNA of *Pneumococcus*⁷, degrades RNA of bacteriophage R17 (ref. 16), induces reverse mutations in *S. typhimurium* (Table 2), fragments DNA of cultured human cells (Fig. 1) and mouse sarcoma cells¹⁷, triggers a DNA repair synthesis (Fig. 2) and elicits chromosome breaks and exchanges (Table 1), as do most, if not all chemical mutagens. The precise molecular nature of the ultimate mutagenic metabolite of ascorbic acid is unknown. Whether hydroxyl radicals that are generated in the interaction of Cu⁺ with H₂O₂, which is in turn produced by the oxidation-reduction activity of the ascorbic acid-dehydroascorbic acid system¹⁸, are responsible for the observed DNA fragmentation and the genetic changes is difficult to assess at present.

It is also difficult to evaluate the degree of genetic hazard posed by vitamin C, its decomposition products, and its interaction with metal ions to man. The effective dose of ascorbic acid in the presence of oxygen or cupric ions varies from 5×10^{-5} M for single-strand scissions of DNA⁶ and cleavage of PM2 DNA (unpublished results of A. R. Morgan *et al.*) to between 1 and 8×10^{-3} M for the induction of DNA lesions and chromosome aberrations in cultured human fibroblasts. Several potent mutagens, including MNNG, MMS and EMS require also high concentrations to trigger a DNA repair synthesis^{4,5}. Thus, the active dose of ascorbic acid products is only two- to tenfold greater than that of potent mutagens if mammalian cells are considered. In the microbial test system, however, ascorbic acid-Cu²⁺ was only weakly mutagenic as compared with MNNG. A too simple application of these *in vitro* data to man can lead to erroneous conclusions. Catalyses may inhibit ascorbic acid-initiated DNA cleavage within the human body. Metabolic products of ascorbic acid may reach the intranuclear DNA molecules only if present in excessive amounts. The ascorbic acid-metal mixtures may have no effect because of the lack of free cupric or ferric ions within cells. Nevertheless, the potential mutagenic capacity of ascorbic acid products should be taken into consideration when a potential health hazard due to the addition of relatively large quantities of vitamin C to food products which contain nitrosatable compounds is evaluated.

H. F. STICH
J. KARIM
J. KOROPATNICK
L. LO

Department of Medical Genetics and
Cancer Research Centre,
University of British Columbia,
Vancouver, Canada V6T 1W5

Received January 9; accepted February 26, 1976.

- ¹ Ames, B. N., Durston, W. E., Yamasaka, E., and Lee, F. D., *Proc. natn. Acad. Sci. U.S.A.*, **70**, 2281-2283 (1973).
- ² Epstein, S. S., Arnold, E., Andrea, J., Bass, W., and Bishop, Y., *Toxic. appl. Pharmac.*, **23**, 288-325 (1972).
- ³ Wyrobek, A. J., and Bruce, W. R., *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- ⁴ Stich, H. F., Kieser, D., Laishes, B. A., and San, R. H. C., *Proc. tenth Can. Cancer Conf.*, 83-110, (University of Toronto Press, Toronto, 1973).
- ⁵ San, R. H. C., and Stich, H. F., *Int. J. Cancer*, **16**, 284-291 (1975).
- ⁶ Bode, V. C., *J. molec. Biol.*, **26**, 125-129 (1967).
- ⁷ McCarty, M., *J. exp. Med.*, **81**, 507-514 (1945).
- ⁸ *Vitamin C* (edit. by Birch, G. B., and Parker, Y. T.) (Applied Sciences Publishers, London, 1974).
- ⁹ Mirvish, S. S., Wallcave, L., Eagen, M., and Shubik, P., *Science*, **177**, 65-68 (1972).
- ¹⁰ Fiddler, W., Pensabene, J. W., Piotrowski, E. G., Doerr, R. C., and Wasserman, A. E., *J. Ed Sci.*, **38**, 1084 (1973).
- ¹¹ Laishes, B. A., and Stich, H. F., *Biochem. biophys. Res. Commun.*, **52**, 827-833 (1973).
- ¹² Stich, H. F., and San, R. H. C., *Mutat. Res.*, **10**, 389-404 (1970).
- ¹³ Stich, H. F., Stich, W., and San, R. H. C., *Proc. Soc. exp. Biol. Med.*, **142**, 1141-1144 (1973).
- ¹⁴ Stich, H. F., San, R. H. C., Miller, J. A., and Miller, E. C., *Nature new Biol.*, **238**, 9-10 (1972).
- ¹⁵ Nishioka, H., *Mutat. Res.*, **31**, 185-189 (1975).
- ¹⁶ Wong, K., Morgan, A. R., and Paranchyk, W., *Can. J. Biochem.*, **52**, 950-958 (1974).
- ¹⁷ Yamafuji, K., Nakamura, Y., Omura, H., Soeda, T., and Gytoku, K., *Z. Krebsforsch.*, **76**, 1-7 (1971).

Conformational differences in bacterial ribosomal RNAs in non-denaturing conditions

THE conservative nature of bacterial ribosomal RNA (rRNA) is clear from the narrow range of its guanosine-cytosine (GC) content compared with the base composition of whole cell DNA. The majority of bacterial DNAs contain from 35 to 75% GC, whereas the GC content of rRNA is restricted to a narrow range of 50-53% (ref. 1). Indeed, the ribosomal DNA (rDNA) of many bacterial species can be isolated from sheared preparations of whole cell DNA solely on the basis of its GC content. Furthermore, nucleic acid hybridisation has indicated considerable homology between the rRNAs of diverse bacterial species²⁻⁴ and helped provide phylogenetic support for taxonomic relationships. In an investigation of 22 different bacteria Pace and Campbell found that organisms whose rRNA showed closer homology to *Escherichia coli* rRNA, showed less rRNA homology to *Bacillus steatothermophilus* rRNA and vice versa². Both sedimentation and gel electrophoresis studies in non-denaturing conditions have shown very small, if any, migratory differences between either the 23S or 16S rRNA species of different bacterial species^{5,6}. By using a modified buffer system coupled with double labelling, we have found that both the large and small species of rRNA from *Streptococcus faecalis*, *B. subtilis*, and *E. coli* can be all distinguished on the basis of significant differences in their mobility in both composite polyacrylamide-Agarose gels and polyacrylamide gels.

The relevant bacterial species were labelled with ³H-5-uridine or ¹⁴C-2-uridine and the rRNA was extracted and electrophoresed in composite polyacrylamide-Agarose gels as described in the legend to Fig. 1.

Figure 1a shows the clear separation of both 23S and 16S peaks of the rRNAs from *S. faecalis* and *E. coli*. Similarly, good separation was achieved between *S. faecalis* and *B. subtilis* (Fig. 1b), and *E. coli* and *B. subtilis* (Fig. 1c). The 23S and 16S peaks of all species were clearly separated by one or more gel slices. It is not possible to make comparisons between different gels, since the buffer temperature was not controlled during electrophoresis and the runs were made on different days. It is important to note that these differences would not be observable if absorbance measurements were used and thus would be overlooked in electrophoretic studies using unlabelled rRNA.

These migratory differences were confirmed by electrophoresis in similar conditions in 3% polyacrylamide gels. The only exception to the migratory patterns seen in composite gels was that the 16S peaks of *E. coli* and *B. subtilis* coelectrophoresed (Fig. 2).

Electrophoresis of these rRNA species was also carried out in composite polyacrylamide-Agarose gels with a standard internal buffer concentration of 1E which is commonly used in RNA electrophoresis⁶. In these conditions, all the bacterial rRNAs coelectrophoresed with the exception of *B. subtilis* 23S rRNA, which migrated slightly behind the *S. faecalis* and *E. coli* 23S RNA peaks (Fig. 3).

If these migratory differences were to be interpreted according to the method of Loening⁶, where distance migrated is inversely proportional to the logarithmic value of the molecular weight, they would be equivalent to 10% differences in molecular weight. Electrophoretic migration in non-denaturing conditions has, however, been shown to be primarily dependent on conformation based on the partially double-stranded nature of single-stranded RNA in aqueous solution. Temperature, ionic strength, pH, base sequence and base composition of RNA all seem to be factors which determine conformation.

A common buffer used for electrophoresis of RNA is the 1E buffer of Loening, which is at a physiological pH and sodium ion concentration. The pH and ion concentration assure that at room temperature the rRNA in gels assumes a conformation similar or identical to its native conformation. Successful

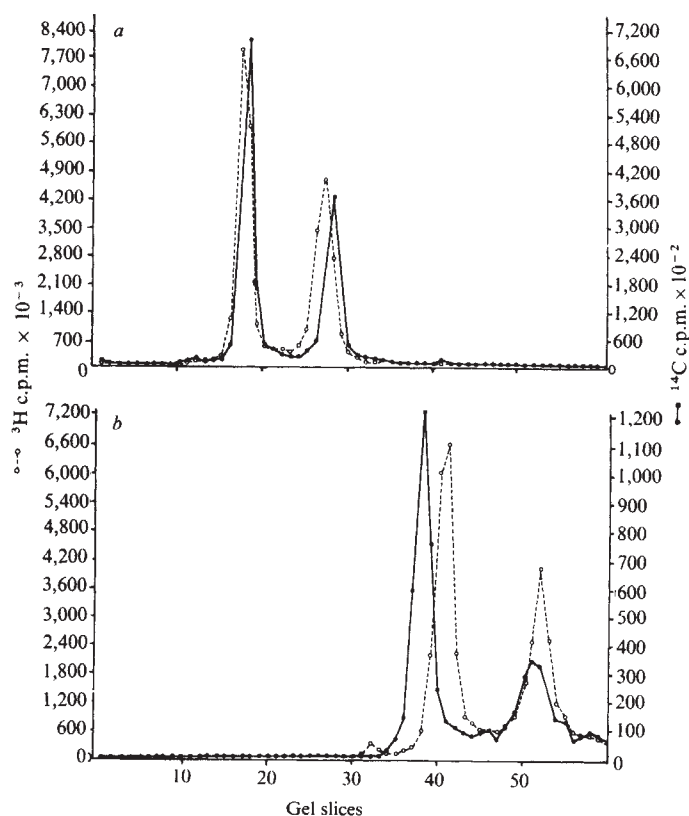


Fig. 1 Bacteria were labelled for 6–12 h of logarithmic growth with either ^3H -5-uridine or ^{14}C -2-uridine (Schwartz-Mann) to facilitate uniform labelling of mature rRNA. RNA was extracted with phenol at 40°C with sodium dodecyl sulphate (SDS)¹¹. All the Gram-positive bacteria were treated before extraction with 0.3% lysozyme for 15 min at 45°C which partially digested the cell wall. RNA was precipitated from the final aqueous phase by addition of cold 0.8 M NaCl and absolute ethanol at -20°C . After incubation overnight at -20°C the mixture was centrifuged and the pellet resuspended in 3 M sodium acetate and 5 mM EDTA to selectively precipitate RNA and remove residual DNA. This solution was centrifuged and the pellet resuspended in a

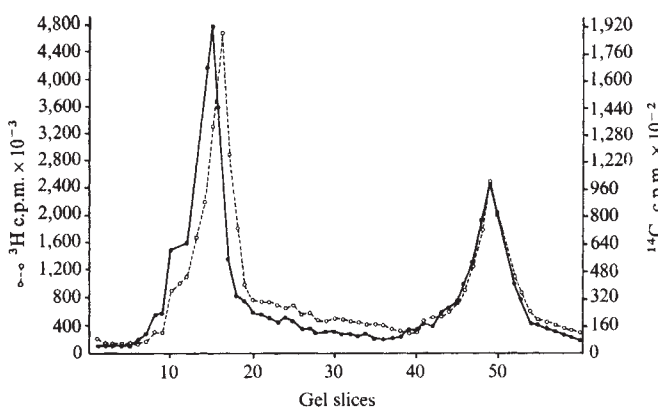
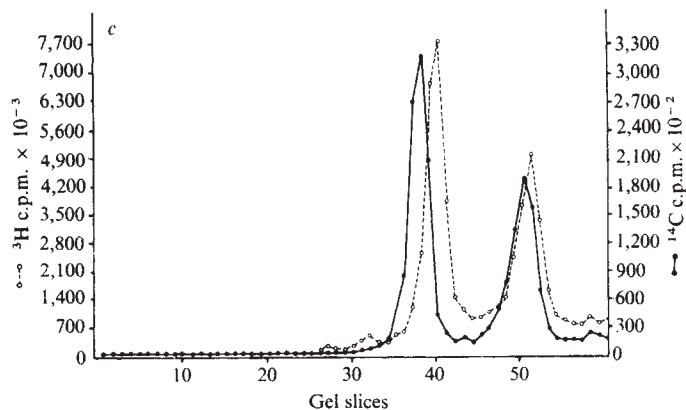


Fig. 2 3% Acrylamide gels were prepared using the monomer mix (acrylamide-bis-acrylamide) of the composite gels. Again, the final concentration of buffer in the gels was 1/3 E. The gels were 14 cm long and 0.6 cm in diameter. Electrophoresis was at 10 mA per gel for 2.5 h. *B. subtilis*, ----; *E. coli*, —.

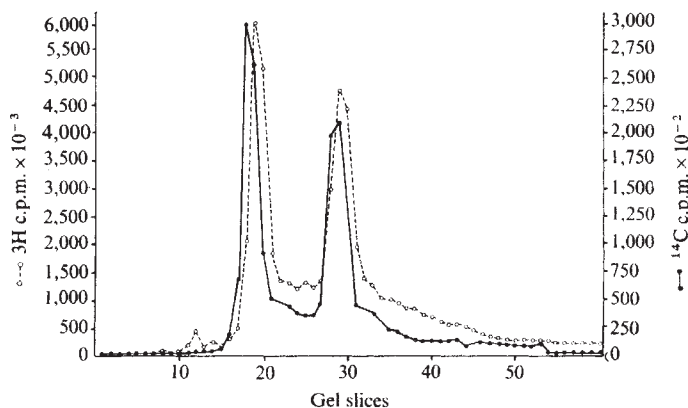
ribosomal reconstitution experiments carried out in similar buffer conditions with homologous rRNA and ribosomal proteins suggests that the rRNA is in a conformation resembling its native state⁷. Successful ribosomal reconstitutions are possible using heterologous rRNA and ribosomal proteins, indicating that the native conformations of rRNAs of different bacterial species have been conserved⁷. It is, therefore, not surprising that

small volume of E buffer plus 0.5% SDS and stored at -70°C (ref. 6). RNA concentration was measured by orcinol determination¹². Generally, 5 μg of each RNA sample in 5–10 μl of E buffer plus 0.5% SDS were added to each gel. RNA samples to be analysed were mixed and placed on the same gel. Preliminary experiments had indicated that the electrophoretic profiles of rRNAs obtained by coextraction did not differ from those obtained by mixing just before the gel run. Polyacrylamide-Agarose gels were prepared by the method of Peacock and Dingman¹³ with the following modifications. The buffer used in gel preparations was 3E rather than 9E so that the final buffer concentration in the gels was 1/3 E rather than 1E¹³. The gels were 2.6% acrylamide, 9–10 cm long and 0.5 cm in diameter. The procedure for gel electrophoresis was that of Loening⁶. The samples were electrophoresed for 1.5 h at 5 mA per gel. In all graphs the direction of electrophoretic migration is from left to right. Because we were interested only in the relative migration of the RNA samples, unwanted sections of the gel were removed from both the top and bottom of the gel. The gel slices referred to on the abscissa therefore do not reflect total distance migrated. After completion of the electrophoretic run the gels were fixed in 1M acetic acid for 15 min, washed with water for at least 30 min, frozen and sliced with a gel slicer (9 slices cm^{-1}). Each slice was placed in a vial with scintillation fluid (PPO-POPOP) + 5% protosol tissue solubiliser, prewarmed to 37°C , and shaken overnight at 37°C . Vials were counted in a Packard TriCarb Liquid Scintillation Spectrometer Model 3003. *a*, *S. faecalis*, ----; *E. coli*, —. *b*, *B. subtilis*, ----; *S. faecalis*, —. *c*, *B. subtilis*, ----; *E. coli*, —.

bacterial rRNAs coelectrophorese in 1E buffer gels.

The differences in migration of both 23S and 16S rRNA seen in Figs 1 and 2 can be attributed to the use of a 1/3E buffer within gels. The lower sodium ion concentration of this buffer may no longer be able to stabilise the secondary and tertiary interactions of segments of the rRNA, thus leading to partial denaturation. The extent of denaturation is probably determined by the GC content of the double-stranded loops and their base

Fig. 3 Composite polyacrylamide-Agarose gels were prepared as in Fig. 1, except the buffer used in gel preparation was 9E, so the final buffer concentration in the gels was 1E. Gels were electrophoresed at 5 mA per gel for 1.5 h. *B. subtilis*, ----; *S. faecalis*, —.



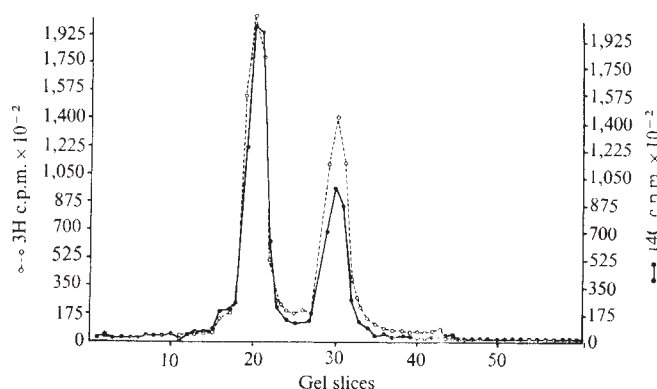


Fig. 4 Formamide gels were prepared according to the method of Staynov *et al.*¹⁴. All other details of the electrophoretic procedures are the same as described in Fig. 1 except the E buffer used in the electrophoresis chambers was adjusted to pH 9.0 rather than 7.8. *B. subtilis*, ----; *E. coli*, —.

sequence. Therefore, rRNAs with a higher GC content presumably retain more stable double-stranded loops, pack tighter, and behave as smaller (lighter) particles during electrophoresis or sedimentation.

We believe that the electrophoretic differences we observed in the 1/3E buffer may be due to small differences in the GC content of the rRNAs of these particular bacterial species. Determinations of their respective GC contents are in progress.

Owing to the extended length of the labelling period, we do not believe any of the species of rRNA described are precursor RNAs. Morris and Schuap have separately reported multiple conformational forms of both 23S and 16S rRNAs from *E. coli*^{8,9}. These forms were only seen, however, when special conditions such as a pH gradient⁸ or electrophoresis for 7.5 h at 0 °C (ref. 9) were used. We have not seen any multiple peaks in our electrophoresis conditions.

To test our assumption that the differences we observe in non-denaturing gels were due primarily to conformational difference, we ran the rRNA samples on formamide gels. Formamide essentially destroys all secondary bonding and reduces migration to a dependency on the primary polynucleotide sequence¹⁰. In all cases we found coelectrophoresis of the different bacterial rRNAs. A single example is shown in Fig. 4. Thus, no molecular weight differences were seen between the respective 23S and 16S rRNA of the different bacterial species studied.

Using a combination of lowered ionic strength in our gels and double labelling of rRNAs we were able to show major migrational differences in the electrophoretic mobility of non-denatured rRNA from various bacterial species. Extension of this work could lead to differential rRNA profiles of bacteria; which might be useful in the identification of non-cultivable pathogens, or to help establish phylogenetic relationships between different bacterial species.

This work was supported by grants from the US National Institutes of Health. We thank Dr Paul Sypherd for helpful discussion.

M. E. REFF
E. J. STANBRIDGE*

Department of Medical Microbiology,
Stanford University School of Medicine,
Stanford, California 94305

Received November 13, 1975; accepted January, 16, 1976.

*Present address: Department of Medical Microbiology, University of California, Irvine, College of Medicine, Irvine, California 92664.

¹ Pace, N. R., *Bact. Rev.*, **37**, 562–603 (1973).

² Pace, B., and Campbell, L. L., *J. Bact.*, **107**, 543–547 (1971).

³ Dubnau, D., Smith, I., Morrell, P., and Marmur, J., *Proc. natn. Acad. Sci. U.S.A.*, **54**, 491–496 (1965).

⁴ Takahashi, H., Saito, H., and Ikeda, Y., *Biochim. biophys. Acta*, **134**, 124–133 (1967).

⁵ Taylor, M. M., Glasgow, J. E., and Storck, R., *Proc. natn. Acad. Sci. U.S.A.*, **57**, 164–169 (1967).

⁶ Loening, U. E., *J. molec. Biol.*, **38**, 355–365 (1968).

⁷ Nomura, M., Traub, P., and Bechmann, H., *Nature*, **219**, 793–799 (1968).

⁸ Schuap, H. W., Best, J. B., and Goodman, A. B., *Nature*, **221**, 864–872 (1969).

⁹ Morris, D. R., Dahlberg, J. E., and Dahlberg, A. E., *Nucleic Acids Res.*, **2**, 447–458 (1974).

¹⁰ Pinder, J. C., Staynov, D. Z., and Gratzner, W. B., *Biochemistry*, **13**, 5367–5373 (1974).

¹¹ Cooper, L. H., and Kay, J. E., *Biochim. biophys. Acta*, **174**, 503–512 (1969).

¹² Shatkin, A. J., *Fundamental Techniques in Virology* (edit. by Habel, K., and Salsman, N. P.), 231–237 (Academic, New York, 1969).

¹³ Peacock, A. C., and Dingman, C. W., *Biochemistry*, **7**, 668 (1968).

¹⁴ Staynov, D. V., Pinder, J. C., and Gratzner, W. B., *Nature new Biol.*, **235**, 108–110 (1972).

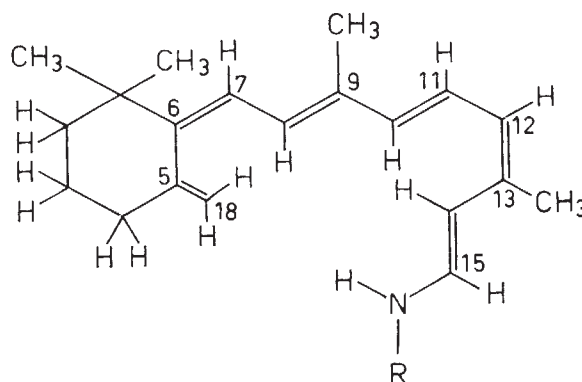
Structure of the chromophoric group in bathorhodopsin

In the photolysis of the visual pigment rhodopsin the intermediate first observed is bathorhodopsin^{1,2} (formerly called prelumi-rhodopsin). It is generally held that isomerisation of 11-*cis* to all-*trans* retinal occurs during this initial step. On theoretical grounds, however, an hexaene-amine structure (Fig. 1) has been proposed as an alternative for the chromophore of bathorhodopsin (ref. 3 and unpublished results of K. Van der Meer, J. J. C. Mulder and J.L.). This chromophore has an exomethylene double bond between the atoms C5 and C18 and the nitrogen of its ene-amine moiety derives from the ε-amino group of a lysine residue of opsin. It is a retrotautomer of the chromophoric group in native rhodopsin which is a protonated Schiff base of 11-*cis* retinal. We now present two lines of evidence which support the structure depicted in Fig. 1: first, the interpretation of a recently published laser resonance Raman spectrum of bathorhodopsin⁴, and second, experiments which establish hydrogen (deuterium) exchange during photolysis of rhodopsin.

The laser resonance Raman spectrum of bathorhodopsin differs from both rhodopsin and isorhodopsin by having additional peak at 1,539, 920, 877 and 856 cm⁻¹. The 1,539 cm⁻¹ absorption is consistent with both the N–H ene-amine structure derived from a primary amine and a protonated Schiff base structure. The 877 cm⁻¹ absorption corresponds with the wagging vibration of 1,1-dialkyl substituted ethene⁵, that of 1,1-dimethylethene is found at 883 cm⁻¹ (ref. 6). Terpenes with an exomethylene bond (for example, sabinene, nopinene and camphene) show three strong peaks in the Raman spectrum around 920, 877 and 856 cm⁻¹ (refs 7 and 8). For comparison we recorded the laser Raman spectrum of vitamin D₂ (in CCl₄), which has a conjugated triene structure with an exomethylene group. The Raman spectrum (Cary 81, He–Ne laser) shows peaks 930, 894, 880 and 863 cm⁻¹. These data strongly suggest that bathorhodopsin, in contrast to native rhodopsin, contains an exomethylene double bond and that the chromophoric group in bathorhodopsin has the structure shown in Fig. 1.

If the retroretinal structure of Fig. 1 were an inter-

Fig. 1 Structure of the chromophoric group in bathorhodopsin.



mediate in the rhodopsin photolysis in the process of reconstituting the 18-CH₃ group in all-*trans* retinal, the ultimate product of rhodopsin photolysis *in vitro*, a hydrogen atom (proton) should be reattached to the 18-CH₂ group in one of the thermal reactions following the formation of bathorhodopsin. Then, hydrogen exchange between chromophore and apoprotein during photolysis might be detected by means of deuteration (or tritiation) of the latter and analysis of the chromophore. In spite of the negative outcome of an earlier attempt to detect tritium exchange during rhodopsin photolysis³, various considerations led us to reinvestigate this approach in a slightly modified manner.

In a first attempt, rhodopsin, isolated as rod outer segments from 60 bovine eyes by a sucrose gradient technique⁹, was washed repeatedly with D₂O and lyophilised. After incubation in D₂O in various conditions, the sample was irradiated with orange light for 30 min. The suspension was brought to 60% (v/v) ethanol and extracted with hexane. The retinal was isolated (thin-layer and high pressure liquid chromatography) in pure form. The samples were placed in an AEI MS-902 mass spectrometer through the direct, heatable introduction system. Spectra were taken, slowly heating the sample from 40 to 120 °C. From the retinal isolated, eight complete mass spectra, each at probe temperatures of 40, 60, 90, 100 and 120 °C, were obtained. At 100 and 120 °C, 20 spectra in the parent peak region were taken. The mass spectra thus obtained closely resemble those published for all-*trans* retinal¹⁰. In three independent experiments in which the incubation conditions with D₂O are varied (0 °C, 14 h; 0 °C, 18 h; 20 °C, 18 h), however, no deuterium incorporation was found in the chromophore.

This result might be explained by the possibility that D₂O has no access to the region of direct protein-chromophore interactions in rhodopsin. Chances for more complete deuteration of the apoprotein could be better in opsin, where the chromophore is absent. Therefore, opsin, isolated as rod outer segments in room light⁹, was washed repeatedly with D₂O and lyophilised. The preparation was resuspended in D₂O and after incubation for various times a fivefold molar excess of 11-*cis* retinal was added. After 1.5 h rhodopsin regeneration, the preparation was centrifuged and the pellet was extracted three times with hexane to remove the excess of retinal. The residue containing regenerated rhodopsin was resuspended in D₂O. Samples of illuminated and unilluminated rhodopsin were extracted, purified and submitted to mass spectrometry as described above. The results are presented in Table 1.

In five independent experiments, deuterium incorporation to the extent of $6 \pm 1\%$ of one D atom was found in the all-*trans* retinal extracted from illuminated samples, as determined from the peaks of $m/e = 284$, 285 and 286, calculated according to Beynon and Williams¹¹. In the control experiments, the chromophore extracted from unilluminated samples (largely 11-*cis* retinal as shown by thin-layer chromatography) did not show any deuterium incorporation. This demonstrates that deuterium is incorporated in the chromophore as a result of illumination.

The accepted mechanism for the photolysis of rhodopsin is based on a photochemical *cis-trans* isomerisation of the chromophoric group¹². Since retinal does not show H-D exchange, in either thermal or photochemical conditions, our observation of a light-induced H-D exchange in the chromophore of rhodopsin is difficult to explain by such a model. We have proposed another mechanism, in which a hydrogen shift is introduced as the photochemical step, resulting in a hexaene-amine moiety as structure for early photointermediates of rhodopsin photolysis (unpublished results of K. Van der Meer, J. J. C. Mulder and J. L.). This model leaves two possibilities for the exchange of hydrogen between chromophore and apoprotein. The

Table 1 Percentage of deuterium incorporation in retinal extracted from rhodopsin obtained by regeneration from D₂O-incubated opsin and 11-*cis* retinal

Opsin incubation with D ₂ O (°C, h)	11- <i>cis</i> Retinal from rhodopsin before illumination	all- <i>trans</i> Retinal from rhodopsin after illumination
20, 1	—	6
20, 5	—	5
20, 14	0	5
20, 18	0	7
30, 14	—	7

hydrogen donated to the apoprotein is either not that which is returned in reconstituting the chromophore or it is later reattached to the chromophore.

In the first case, 100% monodeuteration could be expected if the hydrogen-donating site on the apoprotein is completely deuterated. But our observation that the percentage of deuterium exchange is low and independent of the time during which opsin is incubated with D₂O (Table 1), suggests that such a hydrogen-donating site may not exist. In the second case, one would not necessarily observe H-D exchange; exchange might occur during the time the shuttling hydrogen spends on the protein. We suggest that the very short lifetimes of the photointermediates, bearing a hexaene-amine as chromophore, are responsible for the relatively low deuterium incorporation in retinal during photolysis.

In conclusion, the laser resonance Raman spectrum of bathorhodopsin and the occurrence of hydrogen exchange between chromophore and apoprotein during photolysis support a hydrogen shift as the primary event in the photolysis of rhodopsin.

M. R. FRANSEN

W. C. M. M. LUYTEN

J. VAN THUIJL

J. LUGTENBURG

Laboratory of Organic Chemistry,
University of Leiden,
Leiden

P. A. A. JANSSEN

P. J. G. M. VAN BREUGEL

F. J. M. DAEMEN

Department of Biochemistry,
University of Nijmegen,
Nijmegen, The Netherlands

Received August 4, 1975; accepted February 19, 1976.

¹ Yoshizawa, T., and Kito, Y., *Nature*, **182**, 1604–1605 (1958).

² Busch, G. E., Applebury, M. L., Lamola, A. A., and Rentzepis, P. M., *Proc. natn. Acad. Sci. U.S.A.*, **69**, 2802–2806 (1972).

³ Kropf, A., in *Proc. Int. School of Physics Enrico Fermi*, Course 43, 28–43 (Academic, New York, 1963).

⁴ Oseroff, A. R., and Callender, R. H., *Biochemistry*, **13**, 4243–4248 (1974).

⁵ Sheppard, N., and Simpson, D. M., *Q. Rev., Lond.*, **6**, 1–33 (1952).

⁶ Patnak, C. M., and Fletcher, W. H., *J. molec. Spec.*, **31**, 32–53 (1969).

⁷ Nevgi, G. V., and Jatkari, S. K., *J. Indian Inst. Sci.*, **17A**, 189–196 (1934).

⁸ Dupont, G., Daure, P., and Allard, J., *Bull. Soc. Chim. Fr.*, **49**, 1401–1409 (1931).

⁹ De Grip, W. J., Daemen, F. J. M., and Bonting, S. L., *Vision Res.*, **12**, 1697–1707 (1972).

¹⁰ Lin, R. L., Waller, G. R., Mitchell, E. D., Yang, K. S., and Nelson, E. C., *Analyt. Biochem.*, **35**, 435–441 (1970).

¹¹ Beynon, J. H., and Williams, A. E., *Mass and Abundance Tables for use in Mass Spectrometry* (Elsevier, Amsterdam, 1963).

¹² Wald, G., *Nature*, **219**, 800–807 (1968).

Crystal structure of cholesterol monohydrate

CHOLESTEROL, almost water insoluble, normally occurs in the body bound to lipoprotein, or incorporated in lipid aggregations such as bile micelles, or within certain biomembranes. When cholesterol levels are abnormally high, single crystals of the monohydrate tend to deposit; in bile, clumps of these crystals form gallstones¹. Crystals of cholesterol monohydrate also

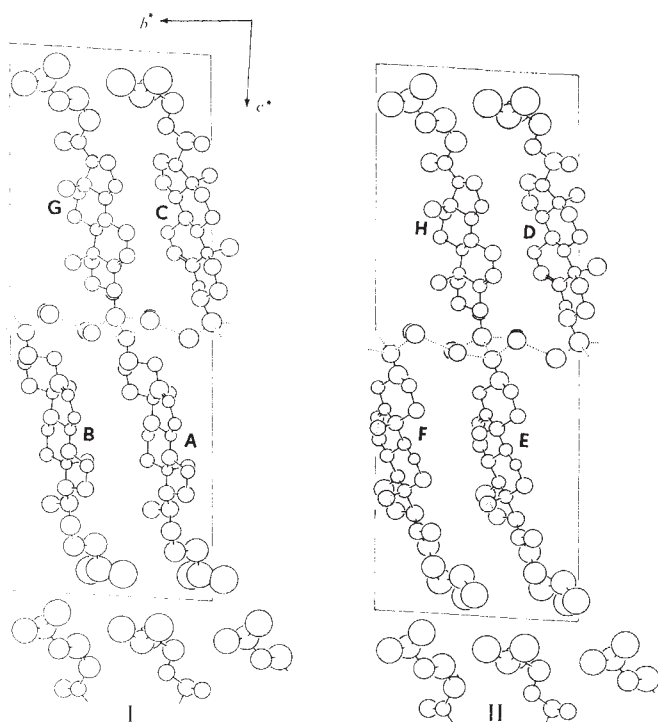


Fig. 1 Crystal structure of cholesterol monohydrate projected down a (into page), with II superposing on I to give the complete structure. Molecules in the pairs (A, B), (E, F) are related by translations $b/2$, and in the pairs (C, D), (G, H) by $a/2$. Circles represent isotropic thermal motion, being drawn at the level of 50% probability of enclosing an atom.

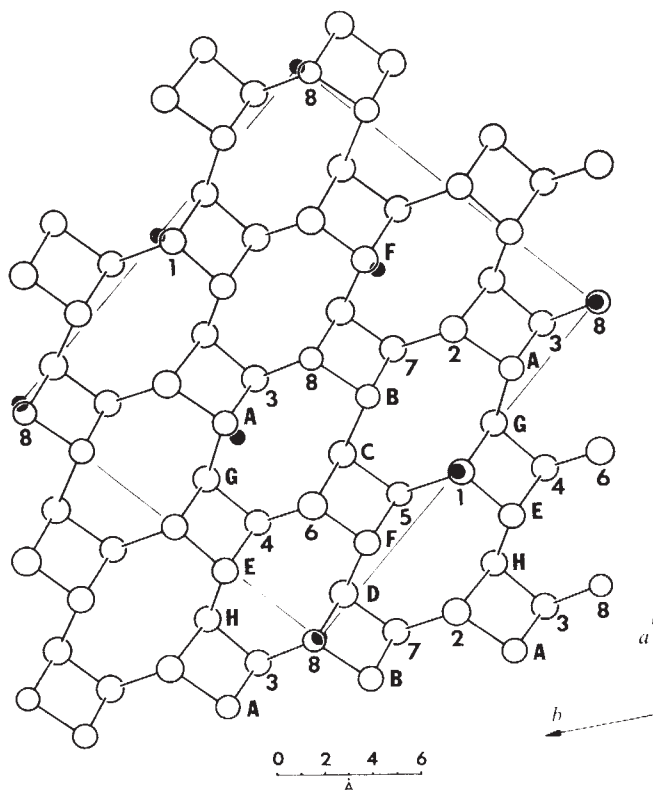
occur in atherosclerotic lesions². Earlier reports of crystal data for cholesterol monohydrate are not all consistent, although they indicate a structure of unusual complexity (refs 1, 3–5 and S. H. Kim and W. Warrant, personal communication). The crystal structure (Fig. 1) has now been determined by a molecular replacement method which was developed primarily for protein crystallography⁶.

A crystal of dimensions $0.4 \times 0.15 \times 0.10$ mm along the a , b , c directions, grown from acetone–water solution and sealed in a lacquer coating, was examined by $\text{CuK}\alpha$ X-ray precession photography and by computer-controlled diffractometer. From comparison of the hkl net with Fig. 2 in ref. 1, the crystalline form appears to be the same as described previously^{1,3}. The diffraction symmetry may well be mistaken for monoclinic, with $(441)^*$ as the approximate twofold axis. There are, however, significant differences in intensities of rows such as $02l$ and $20l$ which reveal the true triclinic symmetry. The space group is thus $P1$, with reduced cell parameters $a=12.39$, $b=12.41$, $c=34.36$ Å, $\alpha=91.9$, $\beta=98.1$, $\gamma=100.8^\circ$. The density (1.045 g cm^{-3}) measured for a crystal from the same mother liquor, and quickly transferred and suspended in a sucrose–water solution, agrees with the calculated density (1.048 g cm^{-3}) assuming the unit cell to contain $8(\text{C}_{27}\text{H}_{46}\text{O}) \cdot 8\text{H}_2\text{O}$ (molecular weight, 3,240). Among the measured 10,157 independent X-ray reflections with $\sin\theta/\lambda < 0.5 \text{ Å}^{-1}$ ($d_{\min}=1 \text{ Å}$) only 2,480 with intensity $I > 2\sigma(I)$ were obtained, and used in the structure determination. The high proportion of weak reflections was due to the large amplitudes of atomic thermal vibrations (overall isotropic $B=7.6 \text{ Å}^2$) and to a remarkable systematic absence of reflections hkl with h and k both odd. These absences were previously assumed to indicate macroscopic twinning on (001) , consistent with the almost equidimensional a and b cell edges¹. It is, however, impossible to resolve the reciprocal lattice reported here into two identical sublattices with a common c^* ,

because the angles α^* and β^* are significantly different (86.5° and 81.4° respectively). It seemed that the twinning must be within the cell unit. A structural model, later shown to be correct, was based on two nearly identical but differently oriented subcells with translations $(a, b/2)$ and $(b/2, a)$. The first subcell structure (for example molecules A, E and B, F, Fig. 1) would contribute no diffracted intensity to reflections hkl with k odd, and the second (for example, molecules C, G and D, H) would contribute nothing to those with h odd, thus explaining the systematic absences. This model was consistent with the maps of the Patterson rotational and translational searches⁷ and was helpful in their interpretation. Eight 20-atom cholesterol fragments were located by comparing the observed Patterson function with that of the same artificial 20-atom structure which was recently used in the crystal structure determination of cholesteryl myristate⁸. All 232 non-hydrogen atoms were found in subsequent electron density maps. Positional and isotropic thermal parameters were refined by a block diagonal least squares procedure to a present R index of 0.12. Estimated standard deviations in bond lengths are about 0.06 Å . Hydrogen atoms were included at calculated positions by assuming C–H bond distances of 1.0 Å and tetrahedral or trigonal bond angles. Hydrogen-bonded hydrogen atoms were not found and thus not included in structure factor calculations. The C(17) chain atoms have large thermal mean square amplitudes of vibration, ranging from 0.25 to 0.40 Å^2 , but there is no evidence of disorder with respect to different conformers.

The structure shows a stacking of bilayers of thickness $d_{001}=33.9 \text{ Å}$. The ring systems of all cholesterols are nearly parallel to (111) with the long axes nearly along $[\bar{1}24]$, which makes an angle of 17.4° with c^* . Molecules such as A and B which are related by subcell translations have the same conformation. Molecules such as A and G which are related by

Fig. 2 A hydrogen-bonded sheet, projected down c^* (into page). Open circles represent thermal parameters for oxygen atoms, as in Fig. 1. Hydroxyl groups are labelled alphabetically and water oxygen atoms numerically. Solid circles represent the hydroxyl ion configuration in the (001) plane of hydroxyapatite. Outlined is a supercell of the hexagonal hydroxyapatite structure.



approximate local twofold symmetry have similar conformations, and molecules such as A and E within the same subcell have different conformations. Thus in A, the C(17) chain is twisted towards the β side of the molecule, whereas in E the twist is towards the α side, so that the two chains are approximately parallel in the crystal. None of the molecules has a fully extended C(17) chain, such as was found for one of the two different rotamers in cholesteryl myristate. The partial folding of all C(17) chains in cholesterol monohydrate results in a condensed structure at the centre of the bilayer.

In projection down $[1\bar{1}24]$, the subcell for cholesterol packing gives a net with $a_s=12.19$, $b_s=6.10$ Å, and $\gamma_s=103.4^\circ$ and an area of 36.2 Å² per molecule. In the bilayers of the crystal structure of cholesteryl myristate⁸, the subcell for cholesterol packing also contains two parallel molecules, the cross-sectional area being 36.6 Å² per molecule. The packing, however, is different. In cholesterol monohydrate, the projecting C(18) and C(19) methyl groups from discrete pairs of molecules (for example A and F) are at van der Waal's distances, whereas in cholesteryl myristate, such distances occur between a given molecule and two neighbouring molecules, thereby forming infinite chains of methyl-methyl interactions.

The C(3) hydroxyl groups and water molecules are hydrogen bonded to form a pleated sheet at the bilayer interface (Fig. 2.) Each oxygen atom forms three hydrogen bonds, so that there are 24 distinct O...O distances which range from 2.68 to 3.01 Å. The average distance (2.86 Å) is on the long side of the reported distribution for other organic hydrates (average 2.80 Å with an e.s.d. of 0.09 Å (ref. 11)). This, together with the rather high values for the mean square thermal amplitudes of vibration of the oxygen atoms ($0.11 < u^2 < 0.16$ Å²), suggests that the hydrogen-bonding network is weak, and that the water molecules have an important space-filling role in stabilising the structure during crystal growth. Migration of water molecules along the tunnels parallel to a , in which they lie (Fig. 1), provides an explanation for the ease with which the crystals dehydrate on exposure to air.

Epitaxial relationships between crystal structures may be important in nucleating the deposit of crystals in biological systems¹⁰. Crystals of both hydroxyapatite and cholesterol monohydrate usually develop (001) faces. The hydroxyl ions of hydroxyapatite are in planes parallel to (001) with O-H bonds favourably oriented for hydrogen bonding at the (001) crystal surface¹¹. A fit of the (001) planes shows close superposition of hydrogen bonding groups from the two crystal structures. The corresponding superlattice nets (Fig. 2) have $a=18.84$, $b=16.31$ Å, $\gamma=90^\circ$ for hydroxyapatite and $a=19.13$, $b=15.84$ Å, and $\gamma=90.1^\circ$ for cholesterol monohydrate. With the intervention of a transition layer of hydrogen-bonded water molecules, a microcrystal of either structure may serve as a nucleus for growth of the other. Since atherosclerotic plaque from recently deceased patients has been shown to contain crystalline hydroxyapatite¹², it is possible that cholesterol monohydrate is implicated in the calcification often associated with later stages in the development of such lesions.

BRYAN M. CRAVEN

Crystallography Department,
University of Pittsburgh,
Pittsburgh, Pennsylvania 15260

Received December 29, 1975; accepted March 10, 1976.

¹ Bogren, H., and Larsson, K., *Biochim. biophys. Acta*, **75**, 65–69 (1963).

² Small, D. M., and Shipley, G. G., *Science*, **185**, 222–229 (1974).

³ Bernal, J. D., Crowfoot, D., and Fankuchen, I., *Proc. R. Soc.*, **A239**, 135–182 (1940).

⁴ Klotzer, F., *Zeit. Kristallogr.*, **95**, 338–367 (1936).

⁵ Tazaki, H., Tatibana, N., and Tatibana, I., *J. Sci. Hiroshima Univ., Ser. III*, **103** (1941).

⁶ Crowther, R. A., in *The Molecular Replacement Method*, (edit. by Rossman, M. G.), 171–177 (Gordon and Breach, New York, 1973).

⁷ Craven, B. M., *ROTRAN Computer Programs* (Technical Report, Crystallography Department, University of Pittsburgh, 1975).

⁸ Craven, B. M., and DeTitta, G. T., *J. chem. Soc.* (in the press).

⁹ Pimental, G. C., and McClellan, A. L., *The Hydrogen Bond* (Freeman, San Francisco, 1960).

¹⁰ Lonsdale, K., *Nature*, **217**, 56–58 (1968).

¹¹ Kay, M. I., Young, R. A., and Posner, A. S., *Nature*, **204**, 1050–1052 (1964).

¹² Spector, M., Krokowsky, E. M., Sax, M., and Pleicher, J., *Science*, **165**, 711 (1969).

Accessible area, packing volumes and interaction surfaces of globular proteins

I RE-EXAMINE here the recently published empirical equations of Chothia¹ for the accessible surface area of 12 monomeric proteins, and demonstrate that accessible surface area of monomeric proteins varies as the $2/3$ power of the molecular weight. The total packing volumes of proteins are obtained from the sum of the individual residues, and the approximation of proteins as spheres is used to predict the loss of accessible surface on association of two proteins. The treatment is extended to calculate the proportion of the monomer surface buried in the formation of higher oligomers.

The total accessible area (in Å²) of the polypeptide chain in the extended form, A_t , was found to be a linear function of molecular weight, M , as it should be since each additional residue contributes on average a fixed area. If A_s is defined as the accessible surface area of the folded protein chain, $A_b = A_t - A_s$ is the accessible surface area that is buried when the protein folds. Chothia proposed that

$$A_t = 1.44 M \quad (1)$$

$$A_b = 0.859 M + 0.774 \times 10^{-5} M^2 \quad (2)$$

It follows that

$$A_s = 0.581 M - 0.774 \times 10^{-5} M^2 \quad (3)$$

Equation 3 implies that the accessible surface is zero at $M = 75,000$, which does not seem reasonable. Moreover, since the proteins are globular it is reasonable to assume that A_s should vary as $M^{2/3}$.

Using Chothia's data for 12 monomeric proteins, I found by least squares analysis that

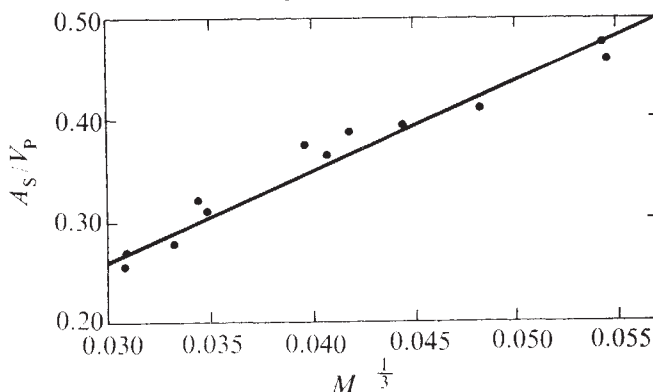
$$A_t = (1.449 \pm 0.006) M \text{ (standard error} = 417 \text{ Å}^2) \quad (4)$$

$$A_s = (11.116 \pm 0.161) M^{2/3} \text{ (standard error} = 401 \text{ Å}^2) \quad (5)$$

The standard error of the fit to the data in equation 5 is not much below the standard error of 428 Å² obtained from the data using equation (3); there are several other reasons, however, that demonstrate that equation 5 has the correct form. First, least squares analysis of $\log A_s$ as a function of $\log M$ yielded 0.685 ± 0.027 as the slope, verifying the $2/3$ exponent for M in equation 5. A second argument for equation 5 comes from the area buried on folding the protein. Subtraction of equation 5 from equation 4 gives

$$A_b = 1.449 M - 11.116 M^{2/3} \quad (6)$$

Fig. 1 Molecular weight dependence of the surface-volume ratio. The solid line corresponds to equation 5. A_s is the accessible surface area and V_p is the total packing volume.



This equation yields zero buried area at $M=451$, a reasonable physical limit. Outside the range of M between 10^4 and 3.5×10^4 equations (2) and (6) differ significantly.

I have estimated the total packing volume of these proteins by simply summing the packing volumes of all protein residues in the primary sequences using the volumes of buried residues given by Chothia¹. The value for the arginyl residue was estimated from the correlation between residue partial specific volumes² and packing volumes for the remaining 19 residues. The values calculated for lysozyme and ribonuclease agree closely with those of Richards³. Defining V_p as the total packing volume, I found the following empirical relationship for the 12 proteins

$$V_p = (1.273 \pm 0.006) M \text{ (standard error} = 401 \text{ \AA}^3) \quad (7)$$

From equation (5) and equation (7), the surface-volume ratio should vary as $8.733 M^{-1/3}$. The least squares equation found is,

$$A_s/V_p = (8.768 \pm 0.102) M^{-1/3} \text{ (standard error} = 0.015 \text{ \AA}^{-1}) \quad (8)$$

Figure 1 shows this surface/volume relation to be reasonable. The exponent of M in equation 8 was verified by analysis of $\log A_s/V_p$ against $\log M$. The exponent found was -0.331 ± 0.021 .

Variation of accessible surface area as $M^{2/3}$ for these monomeric globular proteins implies that in protein-protein complexes, protein subunits may be treated as spheres. In packing volume, monomeric proteins behave as if they possessed the radius,

$$R_p = \left(\frac{3V_p}{4\pi} \right)^{1/3} = \left(\frac{3 \times 1.273}{4\pi} \right)^{1/3} M^{1/3} = 0.6723 M^{1/3} \quad (9)$$

In accessible surface area, the same proteins seem to have radius R_s

$$R_s = \left(\frac{A_s}{4\pi} \right)^{1/2} = \left(\frac{11.116}{4\pi} \right)^{1/2} M^{1/3} = 0.9405 M^{1/3} \quad (10)$$

Thus, for monomeric proteins R_s is $1.4 R_p$. The physical interpretation of this is that the larger value of R_s comes from the roughness of the protein surface.

Chothia and Janin⁴ have demonstrated that the packing volume in the interface between associated proteins is the same as in the proteins themselves. Therefore, on tight protein association, total packing volumes should be additive. On the other hand, some accessible surface becomes buried. Figure 2 shows how this can happen: the total packing volume is conserved in the complex, but the spheres for the surface lose area upon con-

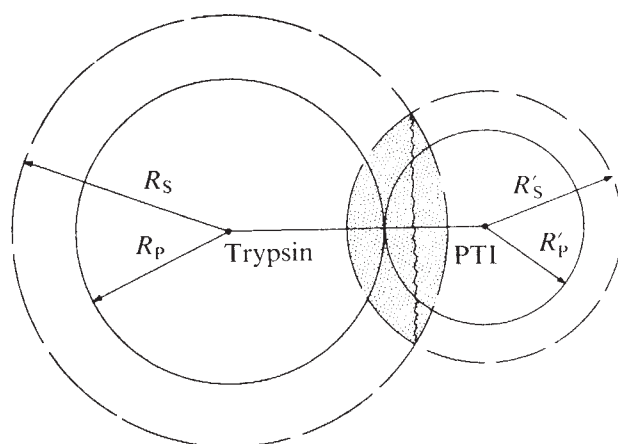


Fig. 2 Schematic diagram of association of trypsin and pancreatic trypsin inhibitor (PTI). The packing radii are represented by R_p and R'_p and the solid circles. The accessible surface radii are denoted by R_s and R'_s and the dashed circles. Accessible surface area which is buried on association is denoted by the overlap of the dashed circles (stippled area). The wavy line denotes the partition of buried accessible surface to the interacting proteins.

tact. To calculate the accessible surface area buried on association, one need only determine the surface area of overlap. Table 1 compares the results of calculations made in this way with those of Chothia and Janin⁴. While not exact, the agreement is remarkably good considering the simplicity of the model. For the dimerisation of identical monomers, the problem can be generally written as, $S_b = 2\pi R_s(R_s - R_p) = 0.143 A_s$. S_b is the surface buried per monomer and A_s is the accessible surface of a monomer as in equation 5. Thus, in general, close association of monomers to form a stable dimer involves a loss of 14.3% of accessible surface area per monomer.

In a square planar tetramer of identical monomers, each monomer contacts two subunits. Thus, each monomer is predicted to lose 28.6% of its accessible surface in the tetramer. The packing of 4 spherical volume elements into a tetrahedral tetramer by this model gives a double burial of some accessible surface area. Neglecting this area of overlap, the geometrical prediction from the model is that 43% of accessible area of the monomeric subunit becomes buried on association with this type of tetramer.

The predicted value of accessible surface area for the concanavalin A monomer from equation 5 is $\sim 6\%$ lower than observed¹. Thus, at present, it is not certain that equation 5 can be used to predict accurately accessible surface of monomers with complex arrangements of subunits in the oligomer.

Table 1 Predicted and observed changes in accessible surface area

Complex	Molecular weight	Total accessible surface area ($\text{\AA}^2$)		Area buried in complex		% Value	
		Predicted	Observed	Predicted	Observed	Predicted	Observed
Insulin dimer							
Monomer M_1	5,787	3,583	3,400	512	530	14	16
Monomer M_2	5,787	3,583	3,310	512	600	14	18
Total change				1,024	1,130		
Trypsin-PTI							
Trypsin	23,302	9,069	8,920	737	640	8	7
PTI	6,158	3,735	3,700	889	815(750)*	24	22(20)*
Total change				1,626	1,455		
Horse haemoglobin							
α chain	15,114	6,795	6,940	991	930	15	13
β chain	15,979	7,052	7,130	983	790	14	11
Total change				1,974	1,720		

*750 $\text{\AA}^2$ was determined from the complex; 815 $\text{\AA}^2$ was found when free PTI coordinates were used.

Nonetheless, the previous considerations indicate that a vast amount of accessible surface becomes buried during the acquisition of quaternary structure. The biological significance of this loss of surface area in relation to oligomeric protein size and structure remains to be established.

I wish to thank Christoph de Haën for discussions, and the National Institutes of Health for support.

DAVID C. TELLER

Department of Biochemistry,
SJ-70, University of Washington,
Seattle, Washington 98195

Received November 13, 1975; accepted March 3, 1976.

¹ Chothia, C., *Nature*, **254**, 304–308 (1975).

² McMeekin, T. L., Groves, M. L., and Hipp, N. J., *J. Am. chem. Soc.*, **71**, 3298–3300 (1949).

³ Richards, F. M., *J. molec. Biol.*, **82**, 1–14, (1974).

⁴ Chothia, C., and Janin, J., *Nature*, **256**, 705–708 (1975).

Increasing effects of repetitive cocaine administration in the rat

REPORTS of chronic cocaine (and related psychomotor stimulant) administration suggest that tolerance develops to many drug effects^{1–3}. A few studies of high dose cocaine administration have suggested that, on the contrary, repetitive administration may be associated with increasing effects on cocaine-induced convulsions in the rat^{4,5} and monkey⁶ and increasing bizarre visual and inhibitory behaviour, as well as dyskinesias, in the monkey⁶. We now report for the first time that repetitive cocaine administration can be associated with progressively increasing effects on horizontal and vertical hyperactivity as well as stereotypy in the rat.

Sixteen male Sprague-Dawley rats weighing 150–185 g were maintained in groups of 3 or 4 animals per cage on a 12-h light-dark cycle. Eight rats received intraperitoneal injections of cocaine 10 mg kg⁻¹ once daily, with two drug-free days each week. Eight saline-control animals received an equal volume of injection on a similar basis. Animals were removed from their home cages and tested in a 24 × 24 × 15 cm Plexiglas chamber. After a 20-min period of adaptation, animals were injected with cocaine or saline, replaced, and monitored in 10-min intervals for 90 min. Each animal was tested at the same time of day, either 0900, 1100, 1300 or 1500, during the light phase of the cycle. Cages were placed on top of 40 infrared photoelectric cells (Motron motility meter) for measurement of activity counts. Slight horizontal movements of the rats, and not just gross locomotor activity, registered as activity counts because of the density of the photocells. Five photoelectric cells positioned 13 cm above the floor of the cage measured vertical activity. Animals were also rated for degree of stereotypy ranging from 0 for absent to +4 for intense, continuous stereotypic patterns during a given 10-min interval.

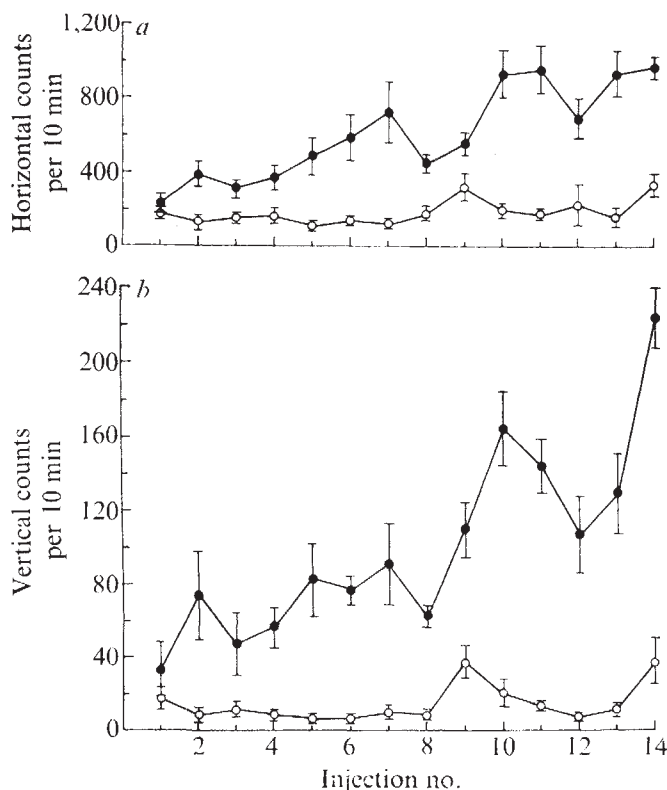
The animals showed increased horizontal and vertical hyperactivity and stereotypy in response to the repeated cocaine injections. Figure 1 illustrates the increase in horizontal and vertical hyperactivity over the course of 14 injections (both $P < 0.001$, 2-way analysis of variance for repeated measures). Figure 2 documents the course of the progressive development of vertical and horizontal hyperactivity, and shows an increase in both peak effects and duration. Following the first few injections, the animals' activity had returned towards the baseline by 60–90 min; while after 15 injections, the activity was still near peak levels at this time. Stereotypy ratings increased from

0.4 ± 0.1 following the first injection to 3.2 ± 0.1 following injection 14. Initially, animals displayed mild repetitive sequences moving from corner to corner, rearing on haunches, and quick circular movements of the head. These patterns were initiated earlier and became progressively more intense, redundant, and prolonged following later injections. For example, rats would engage in frantic, persistent head circling or bobbing while lying prone, or repetitively raise and lower themselves on their haunches with no change of pattern for as long as 15 min.

Conditioning phenomena, as suggested by Tilson and Reich⁷, do not appear to account adequately for these progressive increases in hyperactivity and stereotypy. For injection 25, instead of receiving their usual cocaine injection, experimental animals received saline. As illustrated in Fig. 2, these animals reacted with minimal hyperactivity and stereotypy, not significantly different from saline control animals. Similar results and arguments have been presented by Segal and Mandell⁸ for the progressive effects of continued amphetamine administration on behaviour, again supporting the idea that conditioning effects are not an adequate explanation of the phenomenon. Although increased blood or brain cocaine levels with chronic administration must be carefully ruled out, preliminary evidence in the rat⁹ and monkey (R. Hawks and R. M. P., unpublished) suggests that the peak levels and half life of cocaine in blood and CSF do not increase with chronic administration. After 24 injections, the weights of cocaine treated animals (240 ± 11 g) were similar to that in saline controls (245 ± 9 g), indicating that differential weight loss or debilitation did not account for the progressive effects.

Our data thus suggest that repetitive administration of cocaine is capable of producing increasing hyperactivity and stereotypy in some conditions of administration. Whether tolerance can be manifested may depend not only

Fig. 1 Increasing effect of repeated cocaine injections on (a) horizontal activity and (b) vertical activity. ●, cocaine (10 mg kg⁻¹; n=8); ○, saline (n=8).



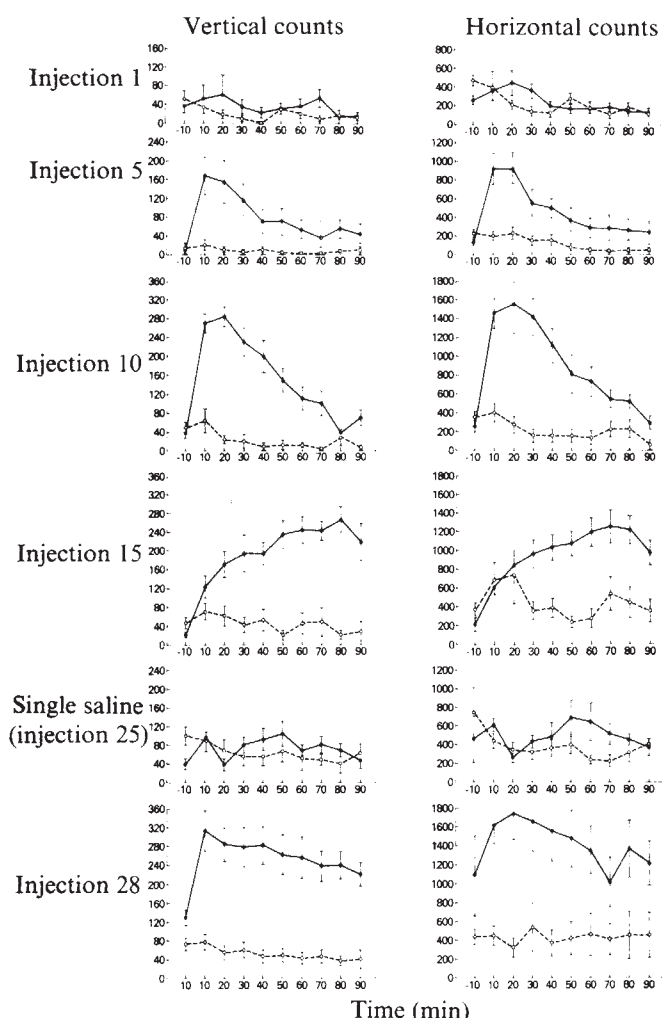


Fig. 2 Effects of repetitive injections on time course of cocaine-induced hyperactivity.

on the specific parameters studied¹⁻³, but also on the intermittent rather than continuous drug administration. Previous studies have documented that repetitive, intermittent, high dose cocaine administration could be associated with progressive increases in a variety of effects in the rat, dog, and monkey, including seizures, catalepsy, motor inhibition, visual tracking behaviour, and dyskinesias^{4,6,10,11}. That the lower doses of cocaine administration used here are also capable of producing progressive effects suggests that the phenomenon is not merely a toxic manifestation of high dose administration.

We have suggested that some progressive effects of repetitive administration of cocaine^{6,12} as well as the local anaesthetic, lidocaine¹³, seem very similar to those of electrical kindling¹⁴⁻¹⁷ in which repetitive electrical stimulation of the limbic system is associated with increasing after-discharge activity and eventually in major motor convulsions at previously subthreshold stimulation. Since cocaine and lidocaine produce marked effects on limbic spindles, after-discharges, and seizure activity^{18,19}, it is possible that a kindling-like mechanism, increasing duration and spread of electrical activity, could account for the progressive effects of lower doses of cocaine where convulsions are not involved as an endpoint. Consistent with a kindling mechanism, Ellinwood *et al.*²⁰ and Stripling and Ellinwood²¹ have reported the increased spread of electrical discharges in cats and rats with chronic cocaine administration.

Many different biochemical mechanisms could also be

involved in the apparent sensitisation to cocaine effects including pre-^{22,23} or post-synaptic^{24,25} alterations, or progressive changes in the activation^{26,27} or binding²⁸ of different kinds of receptors. Rebec and Groves²⁸ reported that repetitive amphetamine pretreatment increased the effect of iontophoretically administered amphetamine and apomorphine on striatal and reticular formation neurones, supporting the concept of an increased receptor sensitivity with chronic stimulant administration²⁵. Gunne and Jonsson³⁰ reported that cocaine injections (100 mg kg⁻¹) once daily for 14 d produced increasing effects on urinary, but not brain, adrenaline and noradrenaline. Study of the biochemical substrates for the progressive behavioural effects of cocaine³¹ and their possible relationship to a kindling process³², may help elucidate the mechanisms involved in the stimulant induced psychoses of man.

ROBERT M. POST
HARVEY ROSE

Adult Psychiatry Branch,
Section on Psychobiology,
NIMH, 9000 Rockville Pike,
Bethesda, Maryland 20014

Received December 15, 1975; accepted March 2, 1976.

- Ritchie, J. M., Cohen, P. J., and Dripps, R. D., in *The Pharmacological Basis of Therapeutics IV* (edit. by Goodman, L. S., Gilman, A.), 371-401 (Macmillan, New York, 1970).
- Caldwell, J., and Seever, P. S., *Clin. Pharmac. Ther.*, **16**, 625-638 (1974).
- Lewander, T., in *Neuropsychopharmacology of Monoamines and Their Regulatory Enzymes* (edit. by Usdin, E.), 221-239 (Raven, New York, 1974).
- Downs, A. W., and Eddy, N. B., *J. Pharmac. exp. Ther.*, **46**, 195-198 (1932).
- Post, R. M., in *Symposium on Contemporary Issues in Stimulant Research* (Plenum, New York, 1976).
- Post, R. M., Kopanda, R. T., and Black, K. E., *Biol. Psychiat.* (in the press).
- Tilson, H. A., and Reich, R. H., *Pharmac. Biochem. Behav.*, **1**, 149-153 (1973).
- Segal, D. S., and Mandell, A. J., *Pharmac. Biochem. Behav.*, **2**, 249-255 (1974).
- Nayak, P. K., Misra, A. L., and Mule, S., *Am. Soc. Pharmac. exp. Ther.*, (1975).
- Gutierrez-Noriega, C., and Zapata-Ortiz, V., *Revista Med. exp.*, **3**, 279-306 (1944).
- Gutierrez-Noriega, C., *Fedn Proc.*, **9**, 280 (1950).
- Post, R. M., and Kopanda, R. T., *Lancet*, **i**, 409-410 (1975).
- Post, R. M., Kopanda, R. T., and Lee, A., *Life Sci.*, **17**, 943-950 (1975).
- Goddard, G. V., McIntyre, D. C., and Leech, C. K., *Expl Neurol.*, **25**, 295-330 (1969).
- Racine, R. J., *Electroenceph. clin. Neurophysiol.*, **32**, 281-294 (1972).
- Wada, J. A., and Sata, M. S., *Neurology*, **24**, 565-574 (1974).
- Pinel, J. P. J., and Van Oot, P. H., *Can. J. Neurosci.*, **2**, 467-475 (1975).
- Eidelberg, E., Lesse, H., and Gault, F. P., in *EEG and Behavior* (edit. Glaser, G. H.), 272-283 (Basic Books, New York, 1963).
- Wagman, I. H., De Jong, R. H., and Prince, D. A., *Archs Neurol.*, **18**, 277-290 (1968).
- Ellinwood, E., Kilbey, M., Castellani, S., and Khoury, C., in *Symposium on Contemporary Issues in Stimulant Research* (Plenum, New York, 1976).
- Stripling, J., and Ellinwood, E., in *Symposium on Contemporary Issues in Stimulant Research* (Plenum, New York, 1976).
- Trendelenburg, V., *Pharmac. Rev.*, **18**, 629-640 (1966).
- Roth, P. H., Walters, J. R., Murrin, L. C., and Morgenroth, V. H., III, in *Pre- and Post-Synaptic Receptors* (edit. Usdin, E., and Bunney, W. E., Jr), 5-48 (Marcel Dekker, Inc., New York, 1975).
- Libet, B., and Tosaka, T., *Proc. natn. Acad. Sci. U.S.A.*, **67**, 667-673 (1970).
- Klawans, H. L., and Margolin, D., *Arch. gen. Psychiat.*, **32**, 725-732 (1975).
- Cools, A. R., Hendricks, G., and Korten, J., *J. Neurol. Trans.*, **36**, 91-105 (1975).
- Costall, B., and Naylor, R., *Eur. J. Pharmacol.*, **33**, 301-312 (1975).
- Creese, I., Burt, D., and Snyder, S., *Life Sci.*, **17**, 993-1002 (1975).
- Rebec, G., and Groves, P., in *Symposium on Contemporary Issues in Stimulant Research* (Plenum, New York, 1976).
- Gunne, L. M., and Jonsson, J., *Psychopharmacologia*, **6**, 125-129 (1964).
- Post, R. M., *Am. J. Psychiat.*, **132**, 225 (1975).
- Post, R. M., and Kopanda, R. T., *Am. J. Psychiat.* (in the press).

Nature Index and Binders

The **Index** for 1975 is now available, price £2.25. Copies of the 1974 index are still on sale, price £3.00. **Binders** for the journal are also available at £7.00 for four (a year of *Nature* fits into four binders).

Postage is included in the above prices. Orders should be sent, accompanied by remittance to Macmillan Journals Ltd, Brunel Road, Basingstoke, Hampshire, England.

reviews

Fascination for discovery

Dorothy C. Hodgkin



The Development of X-Ray Analysis. By the late Sir Lawrence Bragg. Edited by D. C. Phillips and H. Lipson. Pp. viii + 270. (Bell: London, 1975.) £6.50.

THIS book is a joy to read. Over and over again one can almost hear W. L. Bragg talking, as he used to talk when he was with us, always himself fascinated by the problem of solving the next crystal structure, telling us stories meanwhile about the past, mildly debunking, still illuminating.

His point of departure in this book is the discovery of X rays by Röntgen in 1895 and Röntgen's own extensive experiments on their nature. These lead to a general discussion of the evidence for the nature of light and of fundamental particles, and so to the discovery of X-ray diffraction itself, and to the experiments which followed on the structure of crystals and of atoms. In small asides he recaptures moments in the past. "I remember well a fellow student at Jeans' lectures on quantum mechanics, who used to take me aside after the lecture and give me long dissertations on just where Jeans was wrong. This was Bohr—I little knew at the time how great a man was talking to me."

It was one of the important characteristics of W. L. Bragg's own work that he believed in exact and beautiful

experiments. So details are given throughout the book of the actual apparatus used and also of the conditions in the different laboratories in which research was organised. He wrote of the Cavendish laboratory of his time "When I achieved the first X-ray reflections I worked the Rumkorff coil too hard in my excitement and burnt out the platinum contact. Lincoln, the mechanic, was very annoyed, as a contact cost ten shillings, and refused to provide me with another one for a month . . . I could never have exploited my ideas about X-ray diffraction under such conditions." He moved, in fact, to Leeds, to his father's laboratory, where the facilities were on a very different scale, the consequences of W. H. Bragg's own past experience of setting up a physics laboratory of his own in Adelaide.

To solve the first, quite unknown, crystal structures, W. L. Bragg developed direct and simple arguments, leading to unambiguous answers, and based on a thorough understanding, which he illustrates fully, of optical principles and Fourier analysis. The details of structure determination changed a little as each new family of crystal structures was explored—complex ionic compounds, organic compounds, metallic structures, macromolecules. He took great pleasure in the actual structures found, and this is described in a series of chapters, which end with one on imperfect crystallisations on problems for the future. He also took pleasure in using some of the same simple arguments with which he had solved his first structures to guide the analysis of the most complicated ones, particularly of proteins. His one major omission in the book—any treatment of the statistical approach to phase determination now found in use to be very powerful—probably stems from his liking for unambiguous choices. To examine a series of very probable answers did not appeal to him, even if modern computing and his own contributions to structures solved made choice between them easy.

Not surprisingly in a book published posthumously, a few small errors appear, some of which seem sufficiently interesting historically to mention. The first is perhaps obvious—the photograph interpreted by von Laue on p17 is, as mentioned in the text, zinc

Sorry, for copyright reasons some images on this page may not be available online

Dorothy Hodgkin is Wolfson Research Professor of the Royal Society at the Chemical Crystallography Laboratory in the University of Oxford, UK.

blende, not copper sulphate (not interpreted till long afterwards). On p107, the map shown is not of a part of vitamin B₁₂ but of the first synthetic corrin derivative—nice indeed that the electron density here was computed to demonstrate that the correct chemical synthesis had been achieved. On p112, the structure shown in Fig. 4a has a negative peak at the hexagon centre, not a positive peak as described in the text. It illustrates a deduction of Dorothy Wrinch's that the centre of the insulin unit cell should be relatively empty to account, in terms of her theory, for the observed Patterson peak heights. And finally, on p181, Bernal's first indications of the sterol structure were obtained in 1932, not 1943.

This is a book for many different kinds of reader—beginners in X-ray analysis who want to get a feel of the subject, old hands who wish to remember the past, students in other fields who would like to know something of a fascinating region of knowledge and of those who discovered it. We are deeply indebted to Professor Phillips and Professor Lipson for preparing it for publication. □

Vital statistics

Collected Papers of R. A. Fisher. Five Volumes. Edited by J. H. Bennett. (University of Adelaide: Adelaide, 1971-74.) A\$87.00.

THE University of Adelaide has completed its five-volume edition of the scientific papers of the late Sir Ronald Fisher. These preserve the original typography of the nearly 300 items, and make available, in very agreeable form, all Fisher's important publications in scientific journals and some short articles of more general interest. Unlike other great scientists of this century, Fisher has not been widely known to the public, yet few men have so much changed the face of science. In statistical theory, he has profoundly influenced even those who differed from him philosophically. Almost every field of quantitative experimental and observational science has been affected by considerations of experimental design, of randomisation, and of the practice of statistical tests and inference, most of which can be traced to the impact of Fisher's papers and books between 1920 and 1935. To those

who knew Fisher, there is special delight in again savouring the intellectual excitement that the early papers produced and the highly individual style of the author. The argument is often difficult, sometimes obscure, but always strikingly presented. To look again at "Two new properties of mathematical likelihood" is to gain insight into the real nature of statistical inference, so often concealed by the wrappings of rigour in later developments. A series of papers on the rhesus blood factor shows how statistical thought contributed to the unravelling of a genetic complexity—in retrospect simply, but highly controversial at the time. "The logic of inductive inference" remains not only a classic in the development of statistical inference but a superb example of scientific disputation enlivened by human passion. Any library with pretensions to coverage of statistics and statistical genetics needs these volumes, not merely for their historic interest but also for their continuing powers to stimulate new thought, and for the contrast with the pedestrian style and content of so much modern statistical writing. **D. J. Finney**

Control theory

Control Theory in Biology and Experimental Psychology. By F. M. Toates. Pp. 264. (Hutchinson Educational: London, 1975.) £6.50.

ALTHOUGH many biologists, physiologists and psychologists realise that the concepts of control theory provide powerful unifying principles which relate not only the functional behaviour of diverse systems in their own fields, but also the physical systems of engineers and others, there is still great reserve concerning the benefits to be gained from a deep study of the theory. Although a certain amount of reluctance is due to unfamiliarity with the theoretical and numerical methods on which control theory is based, there are even more profound doubts concerning the ultimate usefulness of the subject. Before a biologist embarks on a long and often difficult study, in which he must become familiar with much of the usage of system and control engineers, he must be convinced that the outcome will offer at least one of several possible advantages. He may wish to formally pose problems which are ill-defined in non-numerical terms. Any theoretical or computer studies of a problem must produce results which at least should give reliable guides to the cause-effect relationship in his systems. He should

feel confident to extend the use of any mathematical models developed to regimes beyond the scope of his experimental techniques. Above all, control theory should enable him to sharpen his choice of experiments and to plan a unique and unambiguous route of experimental deduction, for it is in the laboratory that the biologist makes his significant contributions.

There is a wealth of literature on control theory written for the specialist, but much of it is little considered outside the realms of the University. Faced with this, the biological scientist first has a considerable threshold of theoretical technique to overcome and then he must be able to discern which of the abundance of control theoretical techniques are useful to him. There is a distinct shortage of books which are written specifically for the biological sciences and which give a sufficient introduction to the mathematical rigour which will be needed. I have seen texts in which an introduction to control concepts has been attempted deliberately avoiding mathematics but, as Dr Toates knows, this is a sterile approach.

The present book approaches the problem squarely, attempting to show the importance of differential equations, and theoretical and numerical techniques for solving them, in studying the transient behaviour of any system. These methods are firmly set in a back-

ground of biological and psychological processes which, if somewhat basic, are useful enough to demonstrate the power of the techniques in analysis and deduction. The author has clearly concentrated on systems with which he is most familiar and has not tried to cover an extreme range of examples. The result is a neat text which the control engineer would find fascinating reading and which should lead the biologist gently into a more exacting study of this subject. Only when this step has been made will the considerable potential of control theory for the life sciences be realised.

J. M. Nightingale

Space odyssey

Pioneer Odyssey: Encounter with a Giant. By Richard O. Fimmel, William Swindell and Eric Burgess. Pp. 170 (NASA: Washington, 1975.) \$5.50.

THIS is another in the series of commendable little books resulting from the recent (and continuing) exploration of the Solar System by NASA spacecraft. Although the book is in parts self-congratulatory, and reads in places like a publicity handout (which, in a sense, it is), the amount of worthwhile information and the spread of Jupiter pictures from Pioneer provide excellent value for the very modest cost. *Pioneer Odyssey* is highly suitable for use in introductory astronomy courses, but would serve equally well as an introduction to modern planetary science for someone specialising in another discipline but with a peripheral interest in the subject.

A lengthy introductory chapter describes Jupiter, "Giant of the Solar System"—the history of observations of the planet, how its structure probably differs from the terrestrial planets—and gives a thumbnail sketch of ideas concerning the formation of the Solar System. Although the central emphasis of the book is on the technology and techniques of this deep space mission, the section on scientific results from the Jupiter observations (and from study of the interplanetary medium) provides an excellently clear non-technical summary of the Pioneer 10 discoveries. This, and the whole book, is flawed only by the knowledge that Pioneer 11 has since made further investigations which fill out the overall picture rather more fully. I hope that the NASA/Ames team will be persuaded to produce a second edition with a fuller account of the new understanding of Jupiter in the light of Pioneer 11 data. **John Gribbin**

Laser light

Plasmas and Laser Light. By T. P. Hughes. Pp. xxi+518. (Adam Hilger: London, October 1975.) £35.

LASER-related topics now hit the scientific headlines so frequently that it can become quite difficult to distinguish between actual achievement and future plans or aspirations. Confusion of this sort is perhaps inevitable in a field which is interdisciplinary, and which is still growing very rapidly. For example, in the US the relevant ERDA operating budget (supporting research and development administered by the Energy Research and Development Administration), currently totals some \$90 million per annum. Research into laser isotope-separation techniques now constitutes a significant fraction of this budget, but laser fusion still continues to attract the lion's share of the funding. Within the UK a laser facility for fundamental scientific investigations has recently been established at the SRC's Rutherford Laboratory. In view of all the above developments, the recent publication of Dr Hughes' book is particularly opportune.

The historical background, some potential advantages, possible difficulties and competition for the laser-compression approach to controlled thermonuclear fusion have been outlined by me on previous occasions (*Nature*, **239**, 129 (1972); **251**, 99 and **252**, 631 (1974)). These topics constitute the dominant theme of the book, although magnetic containment and plasma diagnostics are also discussed briefly. *Plasmas and Laser Light* is a broad canvas and Dr Hughes has appropriately included an extensive treatment of ionisation and breakdown in gases and solids.

Some other sections of the book provide background information which is readily available elsewhere. In particular, much of the chapters dealing with small amplitude plasma waves, plasma spectroscopy, incoherent scattering and thermonuclear reactions have been the subject of previous specialist monographs. The main interest of this book, however, is its systematic treatment of the interaction of high intensity, coherent, light with solids, gases and plasmas, and its unifying, pedagogical treatment should therefore be of considerable value to students and younger research workers in the field. The discussion of published experimental work is very well illustrated with line diagrams, but is exhaustively comprehensive; thus the reference list is over 50 pages long. Errors in the text are few and trivial (for example, equation 2.87), and the cross referencing is excellent. It is incorrect to suggest (p54) that laser

fusion has introduced a completely new regime of plasma physics; for example, parametric interactions are well known from previous controlled thermonuclear reaction (CTR) work. It is also a pity that there was no discussion of impurity-resonance scattering, nor any pictorial illustration of the growth of hydrodynamic instabilities which have been predicted in two-dimensional laser-compression computer calculations, since these may both have important consequences. Indeed, the degree of spatial symmetry required of alternative laser-compression schemes could well determine the practical feasibility of these approaches to fusion (see Lawrence Livermore Report UCRL-50021-74, p355). With these caveats, the book seems well-balanced, up-to-date, and a pleasure to read. It will certainly become a well-used text in departmental libraries, although a cheaper paperback edition might more realistically suit the pockets of many research students.

I. J. Spalding

Excited states

Excited States in Organic Chemistry. By J. A. Barltrop and J. D. Coyle. Pp. xii+376. (Wiley: London and New York, 1975.) £15.

BEFORE 1965 there were few textbooks in photochemistry but, commencing with the books by Calvert and Pitts and by Turro, the situation has been transformed. Many books are now available and both beginner and expert are well catered for. As a consequence it is necessary for any new book to offer advantages in clarity of presentation, a wide coverage coupled with modernity, appeal to particular groups of chemists, and, not least, a reasonable price, before such a book can be assured of success.

This book is written in a clear, concise style, is adequately cross referenced and develops its material in a systematic fashion. Following an introductory chapter the basic theoretical foundations of the subject are expounded in three chapters (time-independent properties, time-dependent properties and quenching of excited states). The methods specific to photochemistry are then dealt with in detail although the beginner would need to read other books to get a complete presentation of available techniques. A condensed treatment of orbital symmetry and photochemistry follows. The concluding five chapters deal with different chromophores—C=O, C=C, aromatic nitrogen and saturated—the first three of which present the subject in considerable detail.

Both beginner and expert will profit from reading this authoritative work. The good undergraduate interested in photochemistry would grasp the scope of the subject although he would be more likely to purchase and use one of the smaller and cheaper books which give less detail and a more general coverage. The research student in photochemistry would find the book of considerable use and obtain a perspective for approaching the voluminous literature appearing annually in the Specialist Periodical Reports on Photochemistry. Experienced photochemists including inorganic and physical chemists would benefit from studying this book. It is not as encyclopaedic as Calvert and Pitts but this disadvantage is offset by being up-to-date, at least half of the basic material referenced having appeared since 1966; and in the later chapters dealing with the functional groups three-quarters of the material is new. The misprints are few and almost entirely trivial, with the exception of page 163, where electrocyclic is misprinted as electrolytic. The price is prohibitive for a student, and unfortunately, will limit sales amongst the more specialist audience who will benefit most from reading this book.

B. G. Gowenlock

BOOKS

ON PURE
AND APPLIED SCIENCE

Books reviewed or mentioned in this journal are available from stock.

Catalogues on application.

Please state interests.

SCIENTIFIC LIBRARY

ANNUAL SUBSCRIPTION from £5.00

Reduced rates for multiple subscriptions

Available in U.K. only

Prospectus free on request

H. K. LEWIS & Co. Ltd.

LONDON: 136 GOWER STREET,
WC1E 6BS

Telephone: 01-387 4282

Timely review

Point Defects in Solids. Volume 2: Semiconductors and Molecular Crystals. Edited by James H. Crawford, Jr, and Lawrence M. Slifkin. Pp. xvi+480. (Plenum: New York and London, 1975.) \$45.

THE ever-increasing number of published papers on defects in semiconductors is now vast and so it is invaluable to have timely reviews available for the purpose of co-ordination and also to act as starting points for newcomers to the field.

The present volume is in reality a collection of six such review articles; these may be subdivided into two groups. The last chapter by Chadwick and Sherwood is concerned primarily with diffusion processes in molecular crystals, a topic which is obviously still in its infancy. In the penultimate chapter by Walton, the effect of defects on macroscopic properties, such as thermal conductivity of various materials is discussed: this is a specialised topic and in this sense is distinct from the earlier chapters.

The first four chapters deal primarily with the effects of radiation on

crystals and contain discussions of defect creation, the structure of defects and their effects on the optical and electrical properties of the materials; there is also a chapter on impurity and self-diffusion which are of course not unrelated topics. The styles of these articles vary enormously and their contents also reflect the interests and points of view of the authors. The first by Corbett and Bourgoin is a mixture of sensible speculation, which should provoke further work, and a very comprehensive cataloguing of data, with over 400 references cited. By contrast, chapter four by Watkins on electron spin resonance is completely conservative in its approach, which is perhaps a pity, but its appeal to basic physical ideas rather than to mathematics makes it a superb addition to the literature. The other chapters by Casey and Pearson and by Curtis are of intermediate style, but nevertheless are well written and complete the area being covered.

This volume, although inevitably already somewhat dated, is most welcome, but because of its high price will probably be restricted mainly to libraries.

R. C. Newman

Gauge field theory

Gauge Theories of Weak Interactions. (Cambridge Monographs on Mathematical Physics). By J. C. Taylor. Pp. xv+167. (Cambridge University: Cambridge, London and New York, February 1976.) £10.

FOR more than 25 years quantum electrodynamics (QED) has withstood the rigours of time and increasingly sophisticated experiments, and it remains the most successful physical theory yet formulated. Its language is relativistic quantum field theory, and it has always suggested that other interactions of elementary particles should be described in this way. The weak interactions (typical of which is β decay) looked an especially good bet since it was anticipated that a perturbation approach would be even more accurate than QED. All earlier attempts to implement such a programme, however, foundered on the rock of 'renormalisability'. That is to say, the higher order terms in the perturbation series were incalculable because of (unrenormalisable) infinities. The development of the gauge field theories in the last few years has overcome this problem, and at the same time shown how the weak interactions may be unified with QED. This latter feature is necessary in any case, from a practical point of view, since most known weak interactions

involve charged particles which interact with each other electromagnetically whether we like it or not.

It is the view of most particle physicists that this renaissance of field theory is here to stay, and this provides the motivation for John Taylor's excellent book. It is not enough to know that one can in principle calculate higher order effects in such theories, one still has to know how to do it; and this requires mathematical techniques not yet in the repertoire of most practising physicists. The technicalities of the quantisation of gauge fields using path-integral formalism and Fadeev-Popov 'ghosts' are all here. There are chapters on the renormalisation of gauge theories using the Ward-Takahashi identities and dimensional regularisation, together with more familiar material on the construction of the Weinberg-Salam model and other general models. I particularly liked the chapters on spontaneous breaking of global and local symmetries.

All of this is lucidly explained, and, even more astonishingly, in only 150 pages. This book is a 'must' for all theoreticians. I anticipate that the material will soon be in postgraduate courses, so perhaps CUP can bring out a student edition. I can pay the book no higher compliment than to say that it carries on where Bjorken and Drell leave off, and at the same high standard.

D. Bailin

Plant diseases

Plant Pathogenesis. Advanced Series in Agricultural Science, Vol. 2. By Harry Wheeler. Pp. x+106. (Springer-Verlag: Berlin, Heidelberg and New York, 1975). DM39; \$16.

THE Advanced Series in Agricultural Sciences has as a basic objective the integration of theoretical and technical approaches to agriculture. Volume 2 of the Series deals with pathogenesis, described as the sequence of events that occur during the development of a disease, and is based largely on diseases of higher plants caused by fungi and bacteria. A short chapter on concepts and definitions, important for setting the scene in a book of this type, is followed by a review of mechanisms by which pathogens infect, colonise and damage plants; this and other chapters emphasise the early interactions which are likely to be critical in deciding whether or not disease will develop. The third chapter complements the second by describing the reactions of plants to pathogens and does so under the headings structural, functional and metabolic. There follows an account of disease resistance with emphasis on its induction by pathogens and on the role of phytoalexins. Chapter 5 on the Genetics of Pathogenesis deals briefly with genetics of interactions between pathogens and their hosts but also considers at about the same length the biochemistry of the specificity of these interactions. The last chapter of a few pages speculates on the nature of the response of protoplasts to pathogens. There is a good list of about 200 references and a short index.

The author has ranged over much of the field of physiological plant pathology and has presented interesting and thought-provoking assessments of many different problems which should attract much research by plant pathologists and, it is to be hoped, increasingly by scientists from other disciplines. To get the best from this book the reader needs to know a fair amount about the many and various subjects which are reviewed and discussed. This, however, should not in any way deter the many others who will find in this book a very readable and informative account of most of the basic problems in plant pathology that relate to how microorganisms cause disease in plants or, more often, do not do so.

R. K. S. Wood